



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 122955

TO: Elizabeth Kemmerer

Location: 09/937905

Art Unit: 1646

June *2*, 2004 *4090*

Case Serial Number: 09/937905

From: P. Sheppard

Location: Remsen Building

Phone: (571) 272-2529

sheppard@uspto.gov

### Search Notes

371 - 3/31/2000  
JP - 4/1/99

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 1, 2004, 19:32:41 ; Search time 5516 Seconds  
(without alignments)  
6600.462 Million cell updates/sec  
Title: US-09-937-905-1  
Perfect score: 840  
Sequence: 1 gaaccatgtctggtgtca.....aaaaaaaaaaaaaaaaaaaa 840  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	840	100.0	840	6	BD107339	BD107339 Novel cho
2	840	100.0	840	6	BD168026	BD168026 Novel pro
3	840	100.0	840	10	AB028863	AB028863 Mus muscu
4	823.2	98.0	879	10	BC028434	BC028434 Mus muscu
5	598.6	71.3	1183	6	AX235607	AX235607 Sequence
6	597	71.1	1132	9	BC009077	BC009077 Homo sapi
7	595.4	70.9	1135	9	BC017594	BC017594 Homo sapi
8	595.4	70.9	1136	6	BD168027	BD168027 Novel pro
9	595.4	70.9	1190	6	E28764	E28764 HFI2G53 pol
10	593.8	70.7	1199	9	BC008440	BC008440 Homo sapi
11	593.8	70.7	1226	6	BD107338	BD107338 Novel cho
12	593.8	70.7	1226	6	AF132963	AF132963 Homo sapi
13	576.2	68.6	103861	9	AL807761	AL807761 Human DNA
14	576.2	68.6	147844	2	AC068188	AC068188 Homo sapi
15	576.2	68.6	192001	2	AC068376	AC068376 Homo sapi
16	484.6	57.7	3337	9	BC010133	BC010133 Homo sapi
17	484.6	57.7	3337	9	BC010133	BC010133 Homo sapi
18	360	42.9	758	6	E28765	E28765 HFI2G53 pol
19	354	42.1	778	6	BD216184	BD216184 Novel hum
20	348.2	41.5	851	6	BD216494	BD216494 Novel hum
21	323	38.5	409	6	BD215002	BD215002 Novel hum
22	323	38.5	409	6	AR417783	AR417783 Sequence
23	262.8	31.3	913	3	BD113336	BD113336 EST and e
24	248.8	29.6	300	6	AY071553	AY071553 Drosophil
25	245.6	29.2	300	6	BD212079	BD212079 Novel hum
26	218	26.0	1538	3	AK113459	AK113459 Clona int
27	187.8	22.4	642	9	HUMZC34F03	AF088034 Homo sapi
28	178.4	21.2	158450	10	AC068911	AC068911 Mus muscu
29	178.4	21.2	163615	10	AC068913	AC068913 Mus muscu
30	178.4	21.2	233417	10	AL844146	AL844146 Mouse DNA
31	177.8	21.2	32746	2	AC013224	AC013224 Drosophil
32	177.8	21.2	168384	3	AE003493	AE003493 Drosophil
33	177.8	21.2	307761	3	AE003493	AE003493 Rattus no
34	161.2	19.2	250865	2	AC094899	AC094899 Rattus no
35	161.2	19.2	270241	2	AC112064	AC112064 Rattus no
36	161.2	19.2	298446	2	AC117043	AC117043 Rattus no
37	145	17.3	160764	9	AC087661	AC087661 Homo sapi
38	145	17.3	176593	2	AC087661	AC087661 Homo sapi
39	145	17.3	179901	2	AC011006	AC011006 Homo sapi
40	143	17.0	229	6	AR048129	AR048129 Sequence
41	143	17.0	229	6	AR087771	AR087771 Sequence
42	143	17.0	229	6	AR170853	AR170853 Sequence
43	138	16.4	1852	8	AK065321	AK065321 Oryza sat
44	121.8	14.5	121027	2	AL157891	AL157891 Homo sapi
45	121.8	14.5	128228	9	AL138810	AL138810 Human DNA

ALIGNMENTS

RESULT 1  
BD107339  
LOCUS BD107339 840 bp DNA linear PAT 18-SEP-2002  
DEFINITION Novel chondrogenesis promoter.  
ACCESSION BD107339  
VERSION BD107339.1 GI:23202157  
KEYWORDS JP 2002020311-A/2.  
SOURCE Mus sp.  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 840)  
AUTHORS Kato, Y. and Fujimoto, K.  
TITLE Novel chondrogenesis promoter  
JOURNAL Patent: JP 2002020311-A 2 23-JAN-2002;

```
COMMENT
JAPAN SCIENCE AND TECHNOLOGY CORP
OS Mus sp. (mouse)
PN JP 20020203111-A/2
PD 23-JAN-2002
PI 07-JUL-2000 JP 2000206566
PI YUKIO KATO, KATSUMI FUJIMOTO
PC A61K38/00, A61P19/00, A61P19/02, A61P19/08, C07K16/18, C12N15/09,
PC C12Q1/68,
PC G01N33/53, G01N33/566 // C07K14/42, C07K14/51, C07K14/79, A61K37/02,
PC C12N15/00
CC Novel chondrogenesis promoter
FH Key Location/Qualifiers
FT source 1..840
FT /organism="Mus sp. (mouse)"
FEATURES
source
Location/Qualifiers
1..840
/organism="Mus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 100.0%; Score 840; DB 6; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAACCATGTCTGGCTGTCACAGTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 1 GAACCATGTCTGGCTGTCACAGTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG 60
QY 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGGAAACAGTTTACCATCTGGGCT 120
DB 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGGAAACAGTTTACCATCTGGGCT 120
QY 121 GGGTCATGGCAGTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGCTC 180
DB 121 GGGTCATGGCAGTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGCTC 180
QY 181 CCTCAGGCGTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
DB 181 CCTCAGGCGTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
QY 241 AGCAGGACATAAGCGGGCTCCAGCATTAAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 AGCAGGACATAAGCGGGCTCCAGCATTAAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAG 300
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 TTTTCATGAATGCTTATACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 CTGCTGTGATGGCTACCCCTCTGTTTCCAGGACAGAGAGTTTAAATTTACACATCAAGAGA 420
DB 361 CTGCTGTGATGGCTACCCCTCTGTTTCCAGGACAGAGAGTTTAAATTTACACATCAAGAGA 420
QY 421 TGATCAAGAGAAATGAGAAATGTACCTCAGAGAGCTTATACAGATACGATATGTTAG 480
DB 421 TGATCAAGAGAAATGAGAAATGTACCTCAGAGAGCTTATACAGATACGATATGTTAG 480
QY 481 TGGTACCTATTATTGAGAACTCTCTGAAGAGAGAGATCTCAAGAGAGAGAGAGAGAGAG 540
DB 481 TGGTACCTATTATTGAGAACTCTCTGAAGAGAGAGATCTCAAGAGAGAGAGAGAGAGAG 540
QY 541 CCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600
DB 541 CCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600
QY 601 GGGGAGAAACATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 GGGGAGAAACATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 TTGCTGTCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 TTGCTGTCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
```

```
QY 721 GAATTGTGTAAGCCAAGTGGATCCCTAAGCATCTCCAAACAATAAATAAACAACACTCAATTATG 780
DB 721 GAATTGTGTAAGCCAAGTGGATCCCTAAGCATCTCCAAACAATAAATAAACAACACTCAATTATG 780
QY 781 CCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840
DB 781 CCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840
RESULT 2
BD168026 840 bp DNA linear PAT 17-JAN-2003
LOCUS Novel proteins, gene encoding the same and method of utilization
DEFINITION thereof.
ACCESSION BD168026
VERSION BD168026.1 GI:27873838
KEYWORDS WO 0226978-A/1.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 840)
AUTHORS Sha, S., Mukai, H., Aoki, Y. and Nishi, Y.
TITLE Novel proteins, gene encoding the same and method of utilization
JOURNAL Patent: WO 0226978-A 1 04-APR-2002;
JAPAN TOBACCO INC. SHIKEN SHA, HIDEHITO MUKAI, YOSHIKO AOKI, YOSHISUKE
NISHI
COMMENT OS Mouse macrophage cell RAW 264.7
PN WO 0226978-A/1
PD 04-APR-2002
PF 27-SEP-2001 WO 2001JP008446
PR 27-SEP-2000 JP 00P 294191
PI SHIKEN SHA, HIDEHITO MUKAI, YOSHIKO AOKI, YOSHISUKE NISHI PC
C12N15/12, C12N1/21, C07K14/705, C07K16/28, A61K38/00, A61K39/395, PC
A61K45/00,
PC A61K48/00, A61P7/00, A61P43/00, G01N33/15, G01N33/50, PC
G01N33/566//
PC C12P21/08, (C12N15/12, C12R1:91), (C12N1/21, C12R1:19) CC Novel
proteins, gene encoding the same and method of
utilization thereof
FH Key Location/Qualifiers
FT source 1..840
FT /organism="Mouse macrophage cell RAW 264.7".
FEATURES
source
Location/Qualifiers
1..840
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 840; DB 6; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAACCATGTCTGGCTGTCACAGTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 1 GAACCATGTCTGGCTGTCACAGTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG 60
QY 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
DB 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
QY 121 GGGTCATGGCAGTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGCTC 180
DB 121 GGGTCATGGCAGTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGCTC 180
QY 181 CCTCAGGCGTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
DB 181 CCTCAGGCGTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
QY 241 AGCAGGACATAAGCGGGCTCCAGCATTAAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 AGCAGGACATAAGCGGGCTCCAGCATTAAGAGCTGAAAGAGAGAGAGAGAGAGAGAGAG 300
```

```
301 TTTTCATGAATGCTTATACCATGAGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Db
301 TTTTCATGAATGCTTATACCATGAGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Qy
361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCATCAAGAGA 420
Db
361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCATCAAGAGA 420
Qy
421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
Db
421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
Qy
481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
Db
481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
Qy
541 CCATGAATGATGCCAGCTCCTGCGGTTCTGTCGGCGCTCATGGGTGTACGTGT 600
Db
541 CCATGAATGATGCCAGCTCCTGCGGTTCTGTCGGCGCTCATGGGTGTACGTGT 600
Qy
601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Db
601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Qy
661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATG 720
Db
661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATG 720
Qy
721 GAATGTGTAGCCCAAGTGGATGCTTAAGCAATCTCCCAATTAACCAATTAACCAATTAAG 780
Db
721 GAATGTGTAGCCCAAGTGGATGCTTAAGCAATCTCCCAATTAACCAATTAACCAATTAAG 780
Qy
781 CCTTAAATATAAACTCAGCTGCTTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
Db
781 CCTTAAATATAAACTCAGCTGCTTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840

RESULT 3
AB028863
LOCUS AB028863 840 bp mRNA linear ROD 17-JUN-1999
DEFINITION Mus musculus mRNA for MMRP19, complete cds.
ACCESSION AB028863
VERSION AB028863.1 GI:5103282
KEYWORDS MMRP19.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Shi, S., Aoki, Y. and Nishi, Y.
A cDNA sequence from murine monocyte-macrophage
Published Only in Database (1999)
2 (bases 1 to 840)
Shi, S., Aoki, Y. and Nishi, Y.
Direct Submission
Submitted (10-JUN-1999) Shiken Sha, JAPAN TOBACCO INC., Laboratory
of Life Science & Biomolecular Engineering; 6-2, Umeagaka, Aoba-ku,
Yokohama, Kanagawa 227-8512, Japan
(E-mail: shiken_sha@ims.jti.co.jp, Tel: 81-45-972-5901,
Fax: 81-45-972-6205)
Location/Qualifiers
1..840
/organism="Mus musculus"
/mol_type="mRNA"
/sub_species="scares mouse"
/db_xref="taxon:10090"
/clone="MMR19"
/sex="male"
/cell_line="RAW 264.7"
/cell_type="Monocyte-macrophage"
/clone_lib="a cDNA library of RAW 264.7 cell"
1..840
/gene="MMR19"

gene
```

```
6..731
/gene="MMR19"
/codon_start=1
/product="MMRP19"
/protein_id="BAA78906.1"
/db_xref="GI:5103283"
/translation="MSGCOAGDCCSPPCGAQDKHEPRELIPELKQFYHLGWVTGTG
GSLKHGNIYIAPSGVQKRIQPEDMFVCDINEQDISPPASKLKKSQCTPLEWN
AYTMRGAGAVITHSKAAVNIPLFFQGFKITHQEWIKRKTSGYTRYDDMLVY
PIITNPEEKDLKRMAMHANYEPDSCAVLVRHGVYVWGETWEKAKTMCCECYDLED
IAVSMKMGGLDPTQLPVGENGIV"

ORIGIN
Query Match 100.0%; Score 840; DB 10; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAACCATGTCTGGGTGTCAAGGCTCAAGGAGACTGTTCTCGCGCCCTGTGGCGCGCAGG 60
Db 1 GAACCATGTCTGGGTGTCAAGGCTCAAGGAGACTGTTCTCGCGCCCTGTGGCGCGCAGG 60
Qy 61 ACAAGGAGCACCCCGATTCCTGATCCAGAACCTTTGCAAAACAGTTTACCATCTGGGCT 120
Db 61 ACAAGGAGCACCCCGATTCCTGATCCAGAACCTTTGCAAAACAGTTTACCATCTGGGCT 120
Qy 121 GGGTCATCTGGCACTGGAGGGGGAATCAGCTTGAGCATGGCAATCTACATTTGCTC 180
Db 121 GGGTCATCTGGCACTGGAGGGGGAATCAGCTTGAGCATGGCAATCTACATTTGCTC 180
Qy 181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
Db 181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
Qy 241 AGCAGGACATAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCAGTGCTCTC 300
Db 241 AGCAGGACATAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCAGTGCTCTC 300
Qy 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Db 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Qy 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
Db 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
Qy 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
Db 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
Qy 481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
Db 481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
Qy 541 CCATGAATGATGCCAGCTCCTGCGGTTCTGTCGGCGCTCATGGGTGTACGTGT 600
Db 541 CCATGAATGATGCCAGCTCCTGCGGTTCTGTCGGCGCTCATGGGTGTACGTGT 600
Qy 601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Db 601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Qy 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATG 720
Db 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATG 720
Qy 721 GAATGTGTAGCCCAAGTGGATGCTTAAGCAATCTCCCAATTAACCAATTAAG 780
Db 721 GAATGTGTAGCCCAAGTGGATGCTTAAGCAATCTCCCAATTAACCAATTAAG 780
Qy 781 CCTTAAATATAAACTCAGCTGCTTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
Db 781 CCTTAAATATAAACTCAGCTGCTTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
```



## RESULT 4

BC028434

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC028434 879 bp mRNA linear ROD 07-OCT-2003  
IMAGE:1245515), complete cds.

BC028434

BC028434.1 GI:20306407

MGC

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 879)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, D., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2238257

12477932

2 (bases 1 to 879)

Strausberg, R.

Direct Submission

Submitted (23-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [ccapbe-r@mail.nih.gov](mailto:ccapbe-r@mail.nih.gov)

Tissue Procurement: Marcello Bento Soares, Ph.D.

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>

Series: IRAX Plate: 66 Row: 9 Column: 16

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 9790106.

## FEATURES

source

1. .879

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="MGC:41093 IMAGE:1245515"

/tissue\_type="Mammary gland"

/clone\_lib="Soares\_mammary\_gland\_NbMMG"

/lab\_host="DH10B"

/note="Vector: p17T3-Pac"

1. .879

gene

/gene="Mmrip19-pending"  
/note="synonym: Mmrip19"  
/db\_xref="LocusID:56369"  
/db\_xref="MGI:1926788"  
57\_782  
/codon\_start=1  
/product="cDNA sequence AB028863"  
/protein\_id="AAH28434.1"  
/db\_xref="GI:20306408"  
/db\_xref="LocusID:56369"  
/translation="MSGCOAQDCCSRPCGAODKEHPRELIPELCKOFVHLGWVTGTG  
GSLSKHGEIYIAPSGVKERIQPEDMEVCIDINODISEPPASKLKSQCTPLPMN  
ATMKGAGVITHSKAAVMAITLFPQEFKITHQEMIKIRKCTSGGYRVDMLV  
PIITMPEKDKLKERMAHMYPDSCAVLVRHGVYVWGETWEKAKTWCECYDYLFD  
IIVSMKKMDLPTQLPVGENGIV"  
132\_758  
/note="Aldolase II; Region: Class II Aldolase and Adducin  
N-terminal domain. This family includes class II aldolases  
and adducins which have not been ascribed any enzymatic  
function"  
/db\_xref="CDD:pfam00596"

## ORIGIN

Query Match 98.0%; Score 823.2; DB 10; Length 879;  
Best Local Similarity 99.8%; Pred. No. 7.6e-194;  
Matches 825; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GAACCATGTCTGGCTGCTCAAGCTCAAGAGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 60  
DB 52 GAACCATGTCTGGCTGCTCAAGCTCAAGAGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 111  
QY 61 ACAAGAGACACCCCGATTCCTGATCCAGAGACTTTCGAGAACTTTGAGAACTTTTACCATCTGGGCT 120  
DB 112 ACAAGAGACACCCCGATTCCTGATCCAGAGACTTTCGAGAACTTTTACCATCTGGGCT 171  
QY 121 GGGTCACTGCGACTGGAGGGGGAATCAGCTTGAAGCATGCGCATGAATCTACATGCTC 180  
DB 172 GGGTCACTGCGACTGGAGGGGGAATCAGCTTGAAGCATGCGCATGAATCTACATGCTC 231  
QY 181 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 240  
DB 232 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 291  
QY 241 AGCAGACATAAGCGGGCGCTCCAGCATCTAAGAACTGAAAGAGCGAGTGCATCTCCTC 300  
DB 292 AGCAGACATAAGCGGGCGCTCCAGCATCTAAGAACTGAAAGAGCGAGTGCATCTCCTC 351  
QY 301 TTTTTCATGATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360  
DB 352 TTTTTCATGATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 411  
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGAGTGTAAATTTACACATCAAGAGA 420  
DB 412 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGAGTGTAAATTTACACATCAAGAGA 471  
QY 421 TGATCAAGAGATAAGAAATGTACCTCAGGAGCTTATACAGATACGATGATGTTAG 480  
DB 472 TGATCAAGAGATAAGAAATGTACCTCAGGAGCTTATACAGATACGATGATGTTAG 531  
QY 481 TGGTACTTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAGAAAGATGGCTCATG 540  
DB 532 TGGTACTTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAGAAAGATGGCTCATG 591  
QY 541 CCATGAATGATGATACCCAGACTCTCTGTGCGGTTCTTGTGCGGCGTCAATGGGGGTGATGTTG 600  
DB 592 CCATGAACGAGTATACCAGACTCTCTGTGCGGTTCTTGTGCGGCGTCAATGGGGGTGATGTTG 651  
QY 501 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACTCTGTTGACA 660  
DB 652 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACTCTGTTGACA 711  
QY 661 TTGCTGTCTCCATGAAGAAGATGGGACTCCGATCCAAACACAGCTCCCAAGTTGGAGAAATG 720

```

Db      712 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCATCAACACACAGCTCCCGAGTTGGAGAAAATG 771
QY      721 GAATTGTGTAAGCAAGTGGATGCTTAAGCATCTCCAAAGTAATAAAACAAACTCAATTATG 780
Db      772 GAATTGTGTAAGCAAGTGGATGCTTAAGCATCTCCAAAGTAATAAAACAAACTCAATTATG 831
QY      781 CCTTAATAAATACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAA 828
Db      832 CCTTAATAAATACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAA 879

RESULT 5
AX235607
LOCUS   Sequence 28 from Patent WO0164896.
DEFINITION
ACCESSION AX235607
VERSION   AX235607.1 GI:15795627
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Tang,Y.T., Lu,D.A., Bandman,O., Yue,H., Azimzai,Y., Lal,P.,
           Burford,N. and Baughn,M.R.
TITLE     Human enzyme molecules
JOURNAL   Patent: WO 0164896-A 28 07-SEP-2001;
           Incyte Genomics, Inc. (US)
FEATURES
         source
           location/Qualifiers
             1..1183
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"
               /note="Incyte ID No: 2116390CB1"

ORIGIN
Query Match 71.3%; Score 598.6; DB 6; Length 1183;
Best Local Similarity 84.9%; Pred. No. 5.4e-138;
Matches 694; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

QY      4 CCAATGCTGCTGCTCAAGCTC---AAGGACACTGTTCTCGGCGCGTGTGGCGCGAGG 60
Db      92 CCATGCTGCTGCTGCTGCTCGGAGGAGACTGTTGTTCCCGAGATGCGGCGCGAGG 151
QY      61 ACAGGAGACACCCCGATTCTGTATCCAGAACTTTGCCAAACAGTTTACCATCTGGGCT 120
Db      152 ACAGGAGACATCAAGATACCTGATCCAGACTTTGCCAAACAGTTTACCATCTGGGCT 211
QY      121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGCAATGAATCTACATTGCTC 180
Db      212 GGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGCGGATGAATCTACATTGCTC 271
QY      181 CTTGAGCGTGTGAAAAGAGGCGCATTCAGCAGAGACATGTTGTGTGTGACATTAATG 240
Db      272 CTTGAGGAGTGCAAAAGAGACGAATTCAGCCTGGAACATGTTGTGTGTGATATAATG 331
QY      241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTTGAAGAGCTGAAAAGAGCAGTGCACTCCTC 300
Db      332 AAAGGACATAAGTGGACCTTCGCCATCGAAGAGCTTAAAAAGAGCAGTGACTCTCCTC 391
QY      301 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360
Db      392 TTTTCATGAATGCTTACCAATGAGAGAGCAGTGCGAGTGATTCATACCCACTCTAAAG 451
QY      361 CTGCTGTGATGGCTACCGTCTGTTTCCAGACAGGAGTTTAAATTCACATCAAGAGA 420
Db      452 CTGCTGTGATGGCCACCCCTTCTTCTTCAGACGCGGAGTTTAAATTCACATCAAGAGA 511
QY      421 TGATCAAGGATTAAGAAATGTACCTCAGAGAGCTATTACAGATACGATGATGTTAG 480
Db      512 TGATAAAGGATTAAGAAATGTACTCTCCGAGGGGTATTATAGATATGAATGTTAG 571
QY      481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGGATCTCAAAAGAGGATGGCTCATG 540

```

```

Db      572 TGGTACCCATTATTGAGAATACACCTGAGGAGAAAGACCTCAAGATAGAAATGGCTCATG 631
QY      541 CCATGAATCAGTACCCAGACTCTCTGTGCGGTTCCTGTCCGGCGTCTATGGGGTGTACGTGT 600
Db      632 CAATGAATGAATACCCAGACTCTCTGTGAGTACTGTGTGAGAGTATGTGT 691
QY      601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTCTTTGACA 660
Db      692 GGGGGGAAACATGGGAGAAAGCGCAAAACCATGTGTGAGTGTATGACTATTATTATGATA 751
QY      661 TTGCTGTCTCCATGAGAGATGGGACTCGATCCCAACACAGCTCCCGAGTTGGAGAAAATG 720
Db      752 TTGCCGTATCAATGAAGAAGTAGGACTTGAATCTTCACAGCTCCCGAGTTGGAGAAAATG 811
QY      721 GAATTGTGTAAGCCAAAGTGGATGCTTAAGCATCTCCAAACAATAAAAC-AAACCTCAATTAT 779
Db      812 GAATTGTCTAAGCCAAAGAGAGTCTAATTATATACAGAGATAAAGCTAAACGTAATTAT 871
QY      780 GCCTTAATAAATACTCAGCTGCTTTTAAAAAATAAAAAAAAA 816
Db      872 TATTTAATGAAAGCTATTATTTTAAATGAATTGAAA 908

RESULT 6
BC009077 1132 bp mRNA linear PRI 12-NOV-2003
LOCUS   Homo sapiens likely ortholog of mouse monocyte macrophage 19, mRNA
DEFINITION (cDNA clone MGC:9397 IMAGE:3875061), complete cds.
ACCESSION BC009077
VERSION   BC009077.1 GI:14290596
KEYWORDS  MGC.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 1132)
           Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
           Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
           Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
           Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heideh,F.,
           Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
           Scheet,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
           Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
           Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
           McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
           Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
           Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
           Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
           Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
           Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
           Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
           Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
           Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
           Generation and initial analysis of more than 15,000 full-length
           human and mouse cDNA sequences
           human and mouse cDNA sequences
           Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
           22388257
           MEDLINE
           PUBMED 12477932
REFERENCE
AUTHORS   Strausberg,R.
           Direct Submission
           Submitted (30-MAY-2001) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
           NTH-MGC Project URL: http://mgc.nci.nih.gov
           Contact: MGC help desk
           Email: ggapbs-@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu  
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sheed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 13 Row: P Column: 18  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Similarity but not identity to protein.

#### FEATURES

source

1. 1132

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:9397 IMAGE:3875061"

/tissue\_type="Eye, retinoblastoma"

/clone\_lib="NIH\_MGC\_67"

/lab\_host="DH10B"

/notes="vector: pCMV-SPORT6"

1. 1132

/genes="MMRP19"

/db\_xref="LocusID:51074"

42 - 770

/codon\_start=1

/product="CGI-29 protein"

/protein\_id="AAH09077.1"

/db\_xref="GI:14290597"

/db\_xref="LocusID:51074"

/translation="MSGCDAREGDCSSRCGADKEHPRLIPELCKQFVHLGWVGT  
GGISLKHGEIYIARSGVOKERLOPEDMEVCDINEKDLSGSPSKLKKSCQPLFM  
NAYTMKGAVIHTIARSGVOKERLOPEDMEVCDINEKDLSGSPSKLKKSCQPLFM  
VPIENTPEEKIDKRAHAMNPEYDSCAVLVRHGVYWGVTWEKAKTMCEYDYL  
DIASVKKVKGLDPSQLPVGENIV"

misc\_feature

117 - 746

/notes="Aldolase II; Region: Class II Aldolase and Adducin  
N-terminal domain. This family includes class II aldolases  
and adducins which have not been ascribed any enzymatic  
function"

/db\_xref="CDD:pfam00596"

#### ORIGIN

Query Match 71.18; Score 597; DB 9; Length 1132;

Best Local Similarity 84.8; Pred. No. 1.3e-137;

Matches 693; Conservative 0; Mismatches 120; Indels 4; Gaps 2;

QY 4 CCATGCTGGTGTCAAGCTC---AAGGAGACTGTTGTCGGCGCCCTGTGGCGCGCAGG 60  
DB 40 CCATGCTGGTGTGATGCTCGGAGGAGACTGTTGTTCCGGAGATGGCGCGCAGG 99  
QY 61 ACAAGGAGACCCCGGATCTGTATCCAGACTTTCCAAACAGTTTACATCTGGGCT 120  
DB 100 ACAAGGAGACATCAAGATACCTGATCCAGAACTTTGCAACAGCTTTACCATTTAGCT 159  
QY 121 GGGTCACTGGCACTGGAGGGGATCAGCTTGAGCATGGCAATGAATCTACATTCCTC 180  
DB 160 GGGTCACTGGCACTGGAGGGGATGATGCTTGAGCATGGCAATGAATCTACATTCCTC 219  
QY 181 CCTCAGGCGTCAAAAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 240  
DB 220 CTTTCAGAGTCAAAAGAGCGAATTCAGCTTGAGCATGTTTGTGTGTGATATAATG 279  
QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300  
DB 280 AAAAGGACATAAGTGGACCTTCGCCATCGAAGAGCTAAAAAGCCAGTGTACTCCTC 339  
QY 301 TTTTTCATGATGCTTATACCATGAGAGAGCTGGCGCAGTGATTCATACCCACTTAAG 360

Db 340 TTTTTCATGATGCTTACACAATGAGAGGAGCAGGTGCAGTGATTCATACCCACTCTAAAG 399  
QY 361 CTGCTGTGATGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATATACACATCAAGAGA 420  
Db 400 CTGCTGTGATGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATATACACATCAAGAGA 459  
QY 421 TGATCAAAAGGAATAAGGAATGTACCTCAGGAGGCTTATACAGATACGATGATGTTAG 480  
Db 460 TGATAAAAGGAATAAGGAATGTACTTCCGAGGGGTATTATAGATATGATGATGTTAG 519  
QY 481 TGGTACCTTATTGAGACACATCCTGAGAGAGGATCTCAAGAGAAAGGATGGCTCATG 540  
Db 520 TGTATCCCATTTATGAGATACACCTGAGGAGAAAGACCTCAAGATAGAAAGGCTCATG 579  
QY 541 CCATGAATGATGATCCAGACTCTCTGTGCGGTCTTGTCCGGGCTCATGCGGCTGACGTGT 600  
Db 580 CAATGAATGATGATCCAGACTCTCTGTGCACTACTGCTCAGACGTCATGAGTATATGTTGT 639  
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTCTGTTTGACA 660  
Db 640 GGGGGGAAACATGGGAGAAAGGCAAAACCATGTGTGAGTGTATGACTTCTGTTTGAYA 699  
QY 661 TTGCTCTCTCCATGAGAGAGATGGGACTCGATCCACACAGCTCCAGTTGGAGAAATG 720  
Db 700 TTGCCGTATCAATGAGAAAGTAGGACTTGTATCTTCACAGCTCCAGTTGGAGAAATG 759  
QY 721 GAATGTGTGAAGCAAGTGGATGCTTAAAGCATCTCCACAAATAAAAC-AAATCAATTAT 779  
Db 760 GAATGTGTGAAGCAAAAGAAAGTCTAATATATATACAGAGATAAAGCTAAACGTAATTAT 819  
QY 780 GCCTTAATAAATCACTGCTGCTTTTAAAAAATAAA 816  
Db 820 TATTTAAATGAAAGCTATTTTAAAAATGAATGAAA 856  
RESULT 7  
BC017594 1135 bp mRNA linear PRI 03-OCT-2003  
Homo sapiens likely ortholog of mouse monocyte macrophage 19, mRNA  
(cdna clone MGC:714 IMAGE:2988353), complete cds.  
DEFINITION  
BC017594  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1135)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Udwin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abrams,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 1135)  
Strausberg,R.



RESULT 9  
E28764/C

Db	575	TGATAAAGGAATTAAGAAATGTACTTCCGGAGGGTATTATAGATATGATGATATGTTAG 511
QY	481	TGTTACCTATTATTAGAACTCTCTGAAGAGAGAGATCTCAAAGAAAGGATGGCTCATG 540
Db	515	TGGTACCCATTATTAGAAATACACTCGAGAGAAAGACCTCAAAGATAGAATGGCTCATG 456
QY	541	CCATGAATAGTACCCAGACTCCCTGTGCGGGTCTTTGTCCGGGTCATCGGGGTACGTGT 600
Db	455	CAATGGAATGATACCCAGACTCTCTGTGCAGTACTGCTGCAGACGTCATGGAGTATATGTGT 396
QY	601	GGGCGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATATGACTACCTCTGTTTGACA 660
Db	395	GGGGGAAAACATGGGAGAAAGCCAAAACATGTGTGAGTGTATATGACTATTATTATGATA 336
QY	661	TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGACTGGGAGAAAATG 720
Db	335	TTGCGGTATCAATGAAGAAAGTAGGACTTGTATCTTTCACAGCTCCCGACTGGGAGAAAATG 276
QY	721	GAATTGTGTAAAGCAAGTGGATGCTTAAGCATCTCCAAACATAAAAC-AAACTCAATTAT 779
Db	275	GAATTGTCTTAAGCCAAAAGAAAGTCTAATTATATACAGAGATAAAGCTAAACGTAATTAT 216
QY	780	GCCTTAAATAAACTCAGCTGCTTTTAAAAAAGAAA 816
Db	215	TATTAAATGAAGACTATTTTTTAAATGAATTGAAA 179
RESULT 10		
BC008440		
LOCUS		
DEFINITION	BC008440	1199 bp mRNA linear PRI 07-OCT-2000
ACCESSION		Human sapiens likely ortholog of mouse monocytic macrophage 19, mRNA
VERSION		(CDNA clone MGC:14646 IMAGE:4096399), complete cds.
KEYWORDS	BC008440.1	GI:14250070
SOURCE		MGC.
ORGANISM		Homo sapiens (human)
REFERENCE		Homo sapiens
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
		1 (bases 1 to 1199)
		Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Viallon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E., Schneringer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE		Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE		22388257
PUBMED		12477932
REFERENCE		2 (bases 1 to 1199)
AUTHORS		Strausberg, R.
TITLE		Direct Submission
JOURNAL		Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK		NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT		Contact: MGC help desk Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>

Tissue Procurement: ATCC  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdc@axll.stanford.edu](mailto:mdc@axll.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 21 Row: b Column: 5.  
Location/Qualifiers  
1. 1199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:14646 IMAGE:4096399"  
/tissue\_type="Bone marrow, chronic myelogenous leukemia"  
/clone\_lib="NIH MGC 54"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
1. 1199  
/gene="MMP19"  
/db\_xref="LocusID:51074"  
95. .823  
/codon\_start=1  
/product="CGI-29 protein"  
/protein\_id="AAH08440.1"  
/db\_xref="GI:14250071"  
/db\_xref="LocusID:51074"  
/translation="MSGCDAREGDCSRRCGAOKERPYLIPELCKQFVHLGWVTG  
GGGSLKHGDEIYIPSGVQKRIQPEDMFVCDINEKDISGSPSKLKKSQCTPLFM  
NAVYMGAGAVIHTHSSAAVMTLLPGRFKLTHQEMIKGIKKCTGGYRYDDMLV  
VPTIENPEKDLKDRMAVNEYPDSCAVLVRHGGYVYWGVTWEKATMCECYDYLE  
DIASVMKKVGLDPSQLPVGENIV"  
170. .799  
/note="Aldolase II; Region: Class II Aldolase and Adducin  
N-terminal domain. This family includes class II aldolases  
and adducins which have not been ascribed any enzymatic  
function"  
/db\_xref="CDD:pfam00596"

ORIGIN

Query Match	70.7%;	Score 593.8;	DB 9;	Length 1199;
Best Local Similarity	84.6%;	Pred. No. 8.4e-137;		
Matches 691;	Conservative 0;	Mismatches 122;	Indels 4;	Gaps 2;

QY	4	CCATGCTCGCTGCTCAAGTCTCCTGATCCCGAGAACTTTGCAAAACAGTTTATCCATCTCGGCT	60
Db	93	CCATGCTCGCTGATGCTCGGAGGGAGACTGTGTTCCTCCGGAGATGGCGCGCGCAGG	152
QY	61	ACAAGGAGCACCCCGGATTCCTGATCCCGAGAACTTTGCAAAACAGTTTATCCATCTCGGCT	120
Db	153	ACAAGGAGCGTCCAGATACCTGATCCCGAGAACTTTGCAAAACAGTTTATCCATCTAGGCT	212
QY	121	GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGCTC	180
Db	213	GGGTCACTGGGACTGGAGAGGAATTAAGCTTGAAGCATGGCATGAAATCTACATTCGCTC	272
QY	181	CCTCAGCGCTGCAAAAGGAGCGCATTCACGCCAGACATGTTGTGTGTGACATTAATG	240
Db	273	CITTCAGAGTGCAAAAGGAACGAATTCAGCCTGAGACATGTTTGTGTGTGATATAATG	332
QY	241	AGCAGGACATTAAGCGGGCCTCCAGCATCTTAAGAAGCTGAAAAAAGCCAGTGCATCTCCTC	300
Db	333	AAAAGACATAGTGACCTTCGCCATCGAAGAGCTAAAAAAGCCAGTGTACTCCTC	392
QY	301	TTTTTCATGAATGCTTATACCATGAGAGAGCTGGCGAGTGATTCATACCCATCTAAAG	360
Db	393	TTTTTCATGAATGCTTACACAATGAGAGAGCAGTGCAGTATTCATACCCATCTAAAG	452





MEDLINE 20272150  
PUBMED 10810093  
REFERENCE 2 (bases 1 to 1226)  
AUTHORS Lin, W.-C.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAR-1999) Institute of Biomedical Sciences, Academia Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan  
FEATURES  
source Location/Qualifiers  
1..1226  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
78..806  
/codon\_start=1  
/product="CGI-29 protein"  
/protein\_id="AAD27738.1"  
/db\_xref="GI:4680697"  
/translations="MSGCDAGEBDCSRRCAQDKEHPRYLIPELCKQFVHLGHWGTG  
GGGISLKHGDEIVTHSPGVOKERILQPEDMFVDINEKDISGSPSKLKKSQCTPLFM  
NAVTRMGAGAVIHTSKAAVATLPPGRFKITHQEMIKIKKTSGGYRYDDMLV  
VPIENTPEKTLKDRMAHAMNEVPDSCAVLVRHGVYVWGETWEKAKTWCECYDYL  
DIAVSMKKVGLDPSQLPVGNGIV"  
ORIGIN  
Query Match 70.7%; Score 593.8; DB 9; Length 1226;  
Best Local Similarity 84.6%; Pred. No. 8.4e-137;  
Matches 691; Conservative 0; Mismatches 122; Indels 4; Gaps 2;  
QY 4 CCATGTCGTGTCAGCT---CAAGAGACTGTTGCTCGCGCGCTGTGGCGCGCAGG 60  
DB 76 CCATGTCGTGTCAGCTGCGGAGGAGACTGTTGTTCCCGAGATGCGGCGCGCAGG 135  
QY 61 ACAGGAGACCCCGGATCTCTGATCCAGAACTTCCAAACAGTTTACCATCTGGGCT 120  
DB 136 ACAAGGAGCATCCAAAGATACCTGATCCAGAACTTTCGAAACAGTTTACCATTTAGCT 195  
QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATCAATCTACATTGCTC 180  
DB 196 GGGTCACTGGCACTGGAGGAGGAAATAGCTTGAAGCATGGCAATCAATCTACATTGCTC 255  
QY 181 CCTCAGCGGTGCAAAAGAGCGCAATTCAGCAGAGACATGTTTGTGTGTGACATTAATG 240  
DB 256 CTTCAGGAGTGCAAAAGAGCAATTCAGCCTGAAGACATGTTTGTGTGTGATATAATG 315  
QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAGAGAGCTGAAAAGAGCGAGTCACTCTC 300  
DB 316 AAAAGGACATAAGTGGACCTTCGCGCATCGAAGAGCTTAAAAAGCGAGTCACTCTC 375  
QY 301 TTTTCATGAATGCTTATACATGAGAGAGCTGGCGAGTCAATCATACCCACTCTAAAG 360  
DB 376 TTTTCATGAATGCTTACACATGAGAGAGCGAGGAGTCAATCATACCCACTCTAAAG 435  
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCAGACAGGAGTTTAAATTCACATCAAGAGA 420  
DB 436 CTGCTGTGATGGCCACACTTCTCTTTCAGACGCGGAGTTTAAATTCACATCAAGAGA 495  
QY 421 TGATCAAGGAATGAAGAAATGACCTCAGGAGCTATTACAGATACATCATATGTTAG 480  
DB 496 TGATAAAGGAATGAAGAAATGATCTTCGAGGGGTATTAAGATATGATATGTTAG 555  
QY 481 TGGTACCTATTATTGAGAACTCTCTCAGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540  
DB 556 TGGTACCATTTATGAGATACACTGAGGAGAGAGACCTCAAGATAGATGGCTCATG 615  
QY 541 CCATGAATGATACCCAGACTCTGTGCGGTTCTTGTTCGGCGGTCAATGGGTGTACGTGT 600  
DB 616 CAATGAATGAATACCCAGACTCTGTGCGTACTGTTGTCAGACGTCATGGAGTATATGTT 675  
QY 601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660  
DB 676 GGGGGGAAACATGGGAGAGGCGCAAAACCATGTGTGAGTGTATGACTATTATTGTATA 735  
QY 661 TTGCTGTCTCCATGAAGAAGATGGGAGCTCGATCCAAACACAGCTCCCGAGTGGAGAAATG 720

Db 736 TTGCCGTATCAATGAAGAAAGTAGGACTTGATCTTCACAGCTCCCAAGTGGAGAAATG 795  
QY 721 GAATTTGTGAAGCAAGTGCATCCCTAAGCATCCCAACATATAAAC-AAATCAATAT 779  
Db 796 GAATTTGTCTAAGCAAGAAAGTCTAATATATATACAGATTAAGCTAAACGTAATAT 855  
QY 780 GCCTTAATAAAACTCAGCTGCTTTTAAAAAATAAAAA 816  
Db 856 TATTTAAATGAAGCTATTTTAAATGAATTGAAA 892  
RESULT 13  
AL807761 103861 bp DNA linear PRI 08-JAN-2003  
LOCUS Human DNA sequence from clone RP11-196118 on chromosome 9, complete  
DEFINITION sequence.  
ACCESSION AL807761  
VERSION AL807761.8 GI:27645838  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Lloyd, D.  
Direct Submission  
Submitted (08-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquy@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk  
On Jan 11, 2003 this sequence version replaced gi:22416103.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquy@sanger.ac.uk  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-196118 is from the library RP11-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
FEATURES  
source Location/Qualifiers  
1..103861  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-196118"  
/clone\_lib="RP11-11.1"  
ORIGIN



Query Match 68.6%; Score 576.2; DB 9; Length 103861;  
 Best Local Similarity 83.2%; Pred. No. 3.1e-132;  
 Matches 680; Conservative 0; Mismatches 133; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGTCAAGTCT---AAGGAGACTGTGTGTCGGCGCGGTGTGGCGCGAGG 60  
 Db 91885 CCATGCTGGCTGTGAGCGCGCGGAGGAGACTGTGTGTCGGCGAGATGCGGAGCGCAGG 91944

QY 61 ACAAGGAGACCCCGGATCTGTATCCAGAACTTTGCAACAGTTTACCATCTGGGCT 120  
 Db 91945 ACAAGGAGATCCAGATACCTGTATCCAGAACTTTGCAACAGTTTACCATCTGGGCT 92004

QY 121 GGGTCACTGGCTGAGGGGGAATCAGCTTCAAGCATGGCAATCAATCTCATTTGCTC 180  
 Db 92005 GGGTCACTGGGACTGGAGAGGAATGGCTTGAGGCAATGGGGTGAATCTACATTTGCTC 92064

QY 181 CTTGAGGGTGTCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGATTAATG 240  
 Db 92065 CTTGAGGAGTGTCAAAAGGAGCAATTCAGCTTCAAGCAATTAAGGCGAGTGTACTCTC 92124

QY 241 AGCAGACATAGCGGGCTCCAGCATCTCAAGACTGAAAGCTGAAAGGAGCCAGTCCCTC 300  
 Db 92125 AAAGGACATAGTGGACCTTCGCCATCTTAAGAGCTAAATTAAGGCGAGTGTACTCTC 92184

QY 301 TTTTCATGATGCTTATACCATGAGGAGCTGGCGCAGTGTATCATACCCACTCTAAAG 360  
 Db 92185 TTTTCATGATGCTTATACCAATGAGGAGCAGGTGGGTGATTCATACCCACTCTAAAG 92244

QY 361 GTGCTGTGATGCTACCTTCTGTTTCCAGGACAGAGTTTAAATTTACATCAAGAGA 420  
 Db 92245 GTGCTGTGATGCTACCTTCTGTTTCCAGGAGTGGAGTTTAAATTTACCATCAAGAGA 92304

QY 421 TCATCAAGGATATAGGAATCTACTCAGGAGGCTTATACAGATACGATGATGTTAG 480  
 Db 92305 TGATCAAGGATATAGGAATGTACTCTGGAGGGTATTATAGATGATGATGTTAG 92364

QY 481 TGGTACCTTATTATGAGAACACTCTCTGAAGAGAAGATCTCAAGAAAGGATGGCTCATG 540  
 Db 92365 TGGCACCCATTATTGAGATATACCTGAGGAGAGACCTTCAAGATGAGATGGCTCATG 92424

QY 541 CCATGAATGAGTACCAGACTCTGTGGGTTCTGTGCGGCTCATGGGGTGTACGTGT 600  
 Db 92425 CAATGAATGAATACCAGACTCTGTGAGTACTGGTCAGAGCTCATGAGTATATGTGT 92484

QY 601 GGGGAGAACATGGGAGAGCAAAACCATGCTGTGAGTGTATGACTACCTGTTTGA 660  
 Db 92485 GGGGAGAACATGAGAGAGGCAAAACCATGCTGTGAGTGTATGACTATTTTGTGATA 92544

QY 661 TTGCTGTCTCCATGAGAGATGGGACTCGATCCACACAGCTCCAGTTGGAGAAATG 720  
 Db 92545 TTGCGGTATCAATGAAGAAAGTAGACTTGTATCTTCAGCTCCAGTTGGAGAAATG 92604

QY 721 GAATGTGTAAAGCAAGTGGATGCTTACGATCTCAACATATAA-ACAACTCAATAT 779  
 Db 92605 GAATGTGTAAAGCAAGAAAGTCTTAATATATACAGAGATAAAGTTAAACATAATAT 92664

QY 780 GCCTTAAATAAACTCAGCTGCTTTTAAAAA 816  
 Db 92665 TATTTAATGAAGCTATTTTTTTTAAATGAATGAAA 92701

RESULT 14  
 AC068188 147844 bp DNA linear HTG 28-MAY-2000  
 LOCUS Homo sapiens chromosome 8 clone RP11-196118 map 8, WORKING DRAFT  
 DEFINITION SEQUENCE, 20 unordered pieces.  
 AC068188  
 ACCESSION AC068188.2 GI:8099907  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (Bases 1 to 147844)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 8, clone RP11-196118  
 Unpublished  
 2 (Bases 1 to 147844)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,  
 Boguslavsky, L., Boukhgaltier, B., Brown, A., Burkett, G.,  
 Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Gage, D.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., Labrecque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meidram, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 28, 2000 this sequence version replaced gi:7670231.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTHR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9711

Center clone name: 196\_I\_18

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 137250 bases at least Q40

Consensus quality: 142296 bases at least Q30

Consensus quality: 144570 bases at least Q20

Insert size: 157000; agarose-fp

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 20 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 6: contig of 6 bp in length

\* 7 106: gap of 100 bp

\* 107 1338: contig of 1232 bp in length

\* 1339 1438: gap of 100 bp

\* 1439 3068: contig of 1630 bp in length

\* 3169 3168: gap of 100 bp

\* 3169 4802: contig of 1634 bp in length

\* 4803 4902: gap of 100 bp

\* 4903 8244: contig of 3342 bp in length

\* 8245 8344: gap of 100 bp

\* 8345 11578: contig of 3234 bp in length

\*

```
* 11579 11678: gap of 100 bp
* 11679 13891: contig of 2213 bp in length
* 13892 13991: gap of 100 bp
* 13992 15558: contig of 1567 bp in length
* 15559 15658: gap of 100 bp
* 15659 19558: contig of 4000 bp in length
* 19559 19758: gap of 100 bp
* 19759 23937: contig of 4179 bp in length
* 23938 24037: gap of 100 bp
* 24038 28435: contig of 4398 bp in length
* 28436 28535: gap of 100 bp
* 28536 32630: contig of 4095 bp in length
* 32631 38632: contig of 5902 bp in length
* 38633 38732: gap of 100 bp
* 38733 43912: contig of 5180 bp in length
* 43913 44012: gap of 100 bp
* 44013 55699: contig of 11687 bp in length
* 55700 55799: gap of 100 bp
* 55800 70127: contig of 14328 bp in length
* 70128 70227: gap of 100 bp
* 70228 89897: contig of 19570 bp in length
* 89898 89997: gap of 100 bp
* 89998 111024: contig of 21027 bp in length
* 111025 111124: gap of 100 bp
* 111125 129264: contig of 18140 bp in length
* 129265 129364: gap of 100 bp
* 129365 147844: contig of 18480 bp in length.
FEATURES
    source
        1..147844
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="8"
            /map="8"
            /clone="RP11-196118"
            /clone_lib="RP11-196118 Human Male BAC"
        1..6
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right
        107..1338
            /note="assembly_fragment"
            1439..3068
                /note="assembly_fragment"
            3169..4802
                /note="assembly_fragment"
            4903..8244
                /note="assembly_fragment"
            8345..11578
                /note="assembly_fragment"
            11679..13891
                /note="assembly_fragment"
            13992..15558
                /note="assembly_fragment"
            clone_end:SP6
            vector_side:right
        15659..19658
            /note="assembly_fragment"
        19759..23937
            /note="assembly_fragment"
        24038..28435
            /note="assembly_fragment"
        28536..32630
            /note="assembly_fragment"
        32731..38632
            /note="assembly_fragment"
        38733..43912
            /note="assembly_fragment"
        44013..55699
            /note="assembly_fragment"
        55800..70127
            /note="assembly_fragment"
```

```
misc_feature 70228..89897
    /note="assembly_fragment"
misc_feature 89998..111024
    /note="assembly_fragment"
misc_feature 111125..129264
    /note="assembly_fragment"
misc_feature 129365..147844
    /note="assembly_fragment"
ORIGIN
    Query Match 68.6%; Score 576.2; DB 2; Length 147844;
    Best Local Similarity 83.2%; Pred. No. 3.2e-132; Indels 4; Gaps 2;
    Matches 680; Conservative 0; Mismatches 133;
QY 4 CCATGTCTGGCTGTCAAGCTC---AAGGAGAGCTGTGTCTCGCGGCGGTGTGGCGGCAGG 60
DB 127305 CCATGTCTGGCTGTGACGCGCGGGGAGGAGACTGTGTTCCTCCGAGATGCGGAGCGCAGG 127365
QY 61 ACAAGGAGACCCCGGATTCCTGATCCCGAAGAACTTTGCAACAGTTTACCATCTGGGGCT 120
DB 127365 ACAAGGAGCATCCAAGATACCTGATCCCGAAGAACTTTGCAACAGTTTACCATCTGGGGCT 127425
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATCTGCTC 180
DB 127425 GGGTCACTGGCACTGGAGGAGGAATTTGGCTTGAGGCATGGCGGTGAATCTACATCTGCTC 127485
QY 181 CCTCAGGGGTGCAAAAGAGCGCATTCAGCCCAAGAGACATGTTTGTGTGTGACATTAATG 240
DB 127485 CTTTCAGGAGTGCAAAAGGAAACGAATTCAGCTGAGAGACATGTTTGTGTGTGATTAATG 127545
QY 241 AGCAGGACATAAGCGGGCTTCAGCATCTAAGAGCTGAAAAAAGCCAGTCGCTCCTC 300
DB 127545 AAAAGGACATAGTGGACCTTCGCCATCTAAGAGCTAAAAATAAAGCCAGTCGCTCCTC 127605
QY 301 TTTTCATGAATGCTTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360
DB 127605 TTCTCATGAATGCTTTACCATGAGAGGAGCAGTGGCGGTGATTCATACCCACTCTAAAG 127665
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAAAATTACATCAAGAGA 420
DB 127665 CTGCTGTGATGGCCACCCTTCTCTTTTCGAGGATGGAGTTTAAAAATTACCATCAAGAGA 127725
QY 421 TGATCAAGAGGAATAAGGAAATGATCTCCTCAGAGGCTATTACAGATACGATGATATGTTAG 480
DB 127725 TGATCAAGAGGAATAAGGAAATGATCTTCTGAGGGGTATTATAGATATGATGATGTTAG 127785
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAGAAAGATGGCTCATG 540
DB 127785 TGGCACCATTTATTGAGAAATCACCTGAGGAGAAAGACTTCAAGATAGAAATGGCTCATG 127845
QY 541 CCATGAATGAGTACCAGACTCCTGTGGGGTTCCTTGTCCGGCGTCATGGGGTGTAGTGT 600
DB 127845 CAATGAATGAATACCAGACTCCTGTGCAGTACTGTCAGAGCTCATGGAGTATATGTTGT 127905
QY 601 GGGGAGAAACATGGGAGAAAGCAAAAACCATGTGTGAGTGTATTGACTACCTGTTGTTGACA 660
DB 127905 GGGGGGAAACATGAGAGAGGCGCAAAACCATGTGTGAGTGTATTGACTATTTATTGATA 127965
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACAGCTCCAGTTGGGAGAAATG 720
DB 127965 TTGCGGTATCAATGAAGAAAGTAGGACTTTGATCCTTCACAGCTCCAGTTGGGAGAAATG 128025
QY 721 GAATGTGTGAAGCCCAAGTGGATGCCCTTAAGCATCTCCAAACAATAAAA-ACAAACTCAATTAT 779
DB 128025 GAATGTCTAAGCCCAAGAAAGTCTAATTTATATACAGAGATAAAGTTAAACATAATTAT 128085
QY 780 GCGTTAAATAAACTCAGCTGCTTTTAAAAAATAAAAA 816
DB 128085 TATTTAAATGAAGCTATTTTTTTTAAATGAATTGAAA 128122
RESULT 15
AC068376/c
```



```
misc_feature /note="assembly_fragment"
9147..11229
misc_feature /note="assembly_fragment"
11330..13595
misc_feature /note="assembly_fragment"
13696..17546
misc_feature /note="assembly_fragment"
17647..21148
misc_feature /note="assembly_fragment"
21249..25060
misc_feature /note="assembly_fragment"
25161..28967
misc_feature /note="assembly_fragment"
29068..33459
misc_feature /note="assembly_fragment"
33560..36735
misc_feature /note="assembly_fragment
clone_end:SP6
vector_side:right"
36836..43003
misc_feature /note="assembly_fragment"
43104..47964
misc_feature /note="assembly_fragment"
48065..56616
misc_feature /note="assembly_fragment"
56717..63334
misc_feature /note="assembly_fragment"
63435..75034
misc_feature /note="assembly_fragment"
75135..90326
misc_feature /note="assembly_fragment"
90427..102582
misc_feature /note="assembly_fragment"
102683..116335
misc_feature /note="assembly_fragment"
116436..133606
misc_feature /note="assembly_fragment"
133707..159227
misc_feature /note="assembly_fragment"
159328..192001
misc_feature /note="assembly_fragment"

ORIGIN
Query Match 68.6%; Score 576.2; DB 2; Length 192001;
Best Local Similarity 83.2%; Pred.No. 3.3e-132;
Matches 680; Conservative 0; Mismatches 133; Indels 4; Gaps 2;
QY 4 CCATGCTGGCTGTCAGCTC---AAGGAGACTGTGCTCGCGCGCGTGTGGCGGCAGG 60
DB 97034 CCATGCTGGCTGTCAGCGCGCGGAGGAGACTGTTGTTCCCGAGATCGGAGCGCAGG 96975
QY 61 ACAAGGAGCACCCCGATTCCTGATCCAGACTTTTGCAACAGTTTACCATCTGGGCT 120
DB 96974 ACAAGGAGTCACAGATACCTGATCCAGACTTTTGCAACAGTTTACCATCTGGCT 96915
QY 121 GGCTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATTCCTC 180
DB 96914 GGGTCACTGGCACTGGAGGGAATTTGGCTTGAGGCATGGCGGTGAAATCTACATTCCTC 96855
QY 181 CTTGAGGCTGCAAAAGAGCGCATTTAGCCAGAACACATGTTTGTGTGTGACATTAATG 240
DB 96854 CTTGAGGCTGCAAAAGAACGAATTCAGCCTGAGACATGTTTGTGTGTGATATAATG 96795
QY 241 AGCAGGACATAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300
DB 96794 AAAAGGACATAGTGGACCTTCGCCATCTAAGAGCTTAATTAAGCCAGTGTACTCCTC 96735
QY 301 TTTTTCATGAATGCTTTATACATCAGAGGAGCTGGCGCAGTGAATTCATCCACTCTAAAG 360
DB 96734 TTCTCATGAATGCTTACACATCAGAGGAGCAGGTGGGTGATTCATACCACCTCTAAAG 96675
QY 361 CTGCTGTGATGGCTACCCCTTCTCTTCCAGGACAGGAGTTTAAAAATTACACATCAAGAGA 420
|||||
```

```
DB 96674 CTGCTGTGATGGCCACCCTCTCTTTGAGGATGGAGTTTAAANTTACCCATCAAGAGA 96615
QY 421 TGATCAAAAGGAATAAGGAAATGTACTCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
|||||
DB 96614 TGATCAAAAGGAATAAGGAAATGTACTCTCTGGAGGCTATTATAGATATGATGATGTTAG 96555
|||||
QY 481 TGGTACCTTATTATTGAGAACACTCTCTGAGAGAAAGATCTCAAGAAAGGATGGCTCATG 540
|||||
DB 96554 TGGCACCCCATTTATTGAGAATAACACCTGAGGAGAAAGACTTCAAGATAGATGGCTCATG 96495
|||||
QY 541 CCATGAATGAGTACCACGACTCTCTGTGGGTTCTTGTCCGGCTCATGGGTGTTAGTCT 600
|||||
DB 96494 CAATGAATGATACCACGACTCTCTGTGAGTACTGTGTGAGTCTATGGAGTATATGTT 96435
|||||
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACCTGTTTGACA 660
|||||
DB 96434 GGGGGGAAACATGAGAGAAAGGCCAAACCATGTGTGAGTGTATTGACTATTTATTGATA 96375
|||||
QY 661 TTGCTGTCTCCATGAGAAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATG 720
|||||
DB 96374 TTGCGGTATCAATGAGAAAGTAGGACTTGTATCTTCCAGCTCCCGAGTTGGAGAAATG 96315
|||||
QY 721 GAATTTGTGTAAGCCAAAGTGGATGCTTAAGCATCTCCAAACAATAAA-ACAAACTCAATTAT 779
|||||
DB 96314 GAATTTGTGTAAGCCAAAGAAAGTCTTAATTTATACAGATATAAGTTAAACATTAATTAT 96255
|||||
QY 780 GCCTTAAATAAAACTCAGCTGCTTTTAAAAAAGAAAA 816
|||||
DB 96254 TATTTAAATGAAAGCTATTTTAAAAATGAATTGAAA 96218
|||||
```

Search completed: June 2, 2004, 00:00:50  
Job time : 5522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 19:29:51 ; Search time 551 Seconds  
(without alignments)

6476.392 Million cell updates/sec

Title: US-09-937-905-1

Perfect score: 840

Sequence: 1 gaaccatgctggtgtca.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	840	100.0	840	5	AAC61149 DNA encod
2	840	100.0	840	6	ABK12567 DNA encod
3	840	100.0	841	6	ABK47966 Murine cD
4	598.6	71.3	1183	5	Aah75155 Nucleotid
5	595.4	70.9	924	5	AAC61150 DNA encod
6	595.4	70.9	1136	6	ABK47967 Human cDN
7	595.4	70.9	1190	2	Aax05748 Nucleotid
8	593.8	70.7	1226	6	ABK12566 DNA encod
9	546	65.0	602	9	ADB51659 Primary r
10	399.8	47.6	3764	7	ABX63252 Human cDN
11	360	42.9	758	2	Aaz16856 Human gen
12	354	42.1	778	2	Aaz17165 Human gen
13	348.2	41.5	851	2	Aaz15675 Human gen
14	311.4	37.1	426	7	ABZ18972 Group III
15	256	30.5	796	4	ABL17337 Drosophil
16	248.8	29.6	300	2	Aaz12753 Human gen
17	246.4	29.3	479	8	ACH29056 Human adu
18	245.6	29.2	300	2	Aaz12752 Human gen
19	190.4	22.7	664	4	Aah33805 Human col
20	177.8	21.2	2934	4	ABL17336 Drosophil
21	177.8	21.2	10061	4	ABL17116 Drosophil
22	143	17.0	229	2	AAT08638 HPV E6-bi
23	143	17.0	229	2	AAT73916 E6-Bindin

#### ALIGNMENTS

##### RESULT 1

AAC61149  
ID AAC61149 standard; DNA; 840 BP.

XX AC AAC61149;

XX XX

DT 07-FEB-2001 (first entry)

XX XX

DE DNA encoding antigen recognised by Ab capable of inducing G-CSF activity.

XX XX

KW Antigenic protein; antibody; granulocyte colony stimulating factor;

KM G-CSF; cancer therapy; bone marrow suppression; mouse; ds.

XX Mus sp.

OS

PN WO200060075-A1.

XX 12-OCT-2000.

PD

PF 31-MAR-2000; 2000WO-JP002080.

XX

PR 01-APR-1999; 99JP-00095092.

XX

XX (NISB ) JAPAN TOBACCO INC.

PA Sha S, Aoki Y, Nishi Y;

XX

PI WPI; 2001-024452/03.

XX

DR P-PSDB; AAY85635.

XX

PT Gene encoding an antigen recognizing an antibody which induces

XX granulocyte colony stimulating factor (G-CSF) expression for gene therapy

PT and treatment of G-CSF associated disorders e.g. the side effects of

XX cancer therapy.

XX

PS Claim 1; Page 47-49; 58pp; Japanese.

XX

CC The present invention relates to a gene encoding an antigenic protein

CC recognised by an antibody or its fragments which can induce the

CC production of granulocyte colony stimulating factor (G-CSF). Also

CC included in the invention are partial sequences of the gene, antibodies

CC recognising all or part of the antigenic protein, expression vectors

CC containing the gene and host cells transformed by the vector. The gene is

CC used for gene therapy, and compounds identified by screening using the

CC gene sequence are used for treatment and prevention of disorders

CC associated with G-CSF expression such as the side effects of cancer

CC

CC therapy (including bone marrow suppression). The present sequence  
CC represents the murine gene of the invention

XX Sequence 840 BP; 273 A; 174 C; 203 G; 190 T; 0 U; 0 Other;

Query Match 100.0%; Score 840; DB 5; Length 840;

Best Local Similarity 100.0%; Pred. No. 2e-183;

Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAACCATGTCGGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60
DB 1 GAACCATGTCGGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60
QY 61 ACAAGGAGCACCCCGATTCTCTGATCCCAAGACTTTTGCACAGTTTACCATCTGGGCT 120
DB 61 ACAAGGAGCACCCCGATTCTCTGATCCCAAGACTTTTGCACAGTTTACCATCTGGGCT 120
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180
DB 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180
QY 181 CCTCAGCGTGCACAAAGGAGCGATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
DB 181 CCTCAGCGTGCACAAAGGAGCGATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
QY 241 AGCAGGACATAAGCGGGCTCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCATCTCTC 300
DB 241 AGCAGGACATAAGCGGGCTCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCATCTCTC 300
QY 301 TTTTCATGAATGTTATACCATGAGAGAGTGGCGAGTGAATTCATCCCATCTAAAG 360
DB 301 TTTTCATGAATGTTATACCATGAGAGAGTGGCGAGTGAATTCATCCCATCTAAAG 360
QY 361 CTGCTGTGATGCTACCTTCTGTTTCCAGGACAGAGTTTAAATTTACACATCAAGAGA 420
DB 361 CTGCTGTGATGCTACCTTCTGTTTCCAGGACAGAGTTTAAATTTACACATCAAGAGA 420
QY 421 TGATCAAGGAATPAAGGAATGTAACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
DB 421 TGATCAAGGAATPAAGGAATGTAACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
QY 481 TGTGACTTATTAATGAGAACACTCTCTGAAGAGAGGATCTCAAAGAAAGGATGGCTCATG 540
DB 481 TGTGACTTATTAATGAGAACACTCTCTGAAGAGAGGATCTCAAAGAAAGGATGGCTCATG 540
QY 541 CAGTGAATGATGATCCAGACTCTGTCGGGTTCTTTCGGGCTCATGGGTTGACGCTCT 600
DB 541 CAGTGAATGATGATCCAGACTCTGTCGGGTTCTTTCGGGCTCATGGGTTGACGCTCT 600
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTTGTGAGTGTATGACTACCTGTTTGACA 660
DB 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTTGTGAGTGTATGACTACCTGTTTGACA 660
QY 661 TTGCTGCTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGATTGGAGAAAATG 720
DB 661 TTGCTGCTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGATTGGAGAAAATG 720
QY 721 GAATTTGTGAAGCAAGTGAATGCTTCAAGCATCTCAACATTAACCAACTCAATTAATG 780
DB 721 GAATTTGTGAAGCAAGTGAATGCTTCAAGCATCTCAACATTAACCAACTCAATTAATG 780
QY 781 CCTTAATAAATCTCAGTCTCTTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
DB 781 CCTTAATAAATCTCAGTCTCTTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
```

RESULT 2

ABK12567

ID ABK12567 standard; cDNA; 840 BP.

XX

AC ABK12567;

XX

DT 18-JUN-2002 (first entry)

XX

DE

XX

KW

KW

KW

KW

KW

XX

OS

XX

XX

FT

FT

FT

FT

XX

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DNA encoding cell differentiation stimulator associated protein #2.

Cartilage cell differentiation stimulator; osteopathic;

Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;

Membrane-bound type transferrin-like protein; Mtf; cartilage disorder;

bone metabolism disease; cell differentiation; cell growth;

extracellular matrix related disease; gene; ss; mouse.

Mus sp.

Location/Qualifiers

6..731

/\*tag= a

/product= "Cell differentiation stimulator associated

protein #2"

JP2002020311-A.

23-JAN-2002.

07-JUL-2000; 2000JP-00206566.

07-JUL-2000; 2000JP-00206566.

(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

WPI: 2002-287405/33.

P-PSDB; AAU78361.

A cartilage cell differentiation stimulator useful in the diagnosis of

biophysics, cell differentiation, cell growth and construction of

extracellular matrix related diseases.

Claim 5; Page 9; 17pp; Japanese.

The invention describes a cartilage cell differentiation stimulator

(containing a membrane-bound transferrin-like protein (Mtf-BP) and a

membrane bound type transferrin-like protein (Mtf)) and an animal-derived

concanavalin-like drug. The cartilage differentiation stimulator can be

used in diagnosis, prevention and treatment of cartilage and bone

metabolism diseases. They can also be used for diagnosing biophysics,

cell differentiation, cell growth and construction of extracellular

matrix related diseases. Mtf-BP strongly stimulates differentiation of

cartilage cells and exhibits similar action mechanism with that of plant

derived ConA. This sequence represents a cartilage cell differentiation

stimulator associated polypeptide described in the invention

Sequence 840 BP; 273 A; 174 C; 203 G; 190 T; 0 U; 0 Other;

Query Match 100.0%; Score 840; DB 6; Length 840;

Best Local Similarity 100.0%; Pred. No. 2e-183;

Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACCATGTCGGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60

DB 1 GAACCATGTCGGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60

QY 61 ACAAGGAGCACCCCGATTCTCTGATCCCAAGACTTTTGCACAGTTTACCATCTGGGCT 120

DB 61 ACAAGGAGCACCCCGATTCTCTGATCCCAAGACTTTTGCACAGTTTACCATCTGGGCT 120

QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180

DB 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180

QY 181 CCTCAGCGTGCACAAAGGAGCGATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240

DB 181 CCTCAGCGTGCACAAAGGAGCGATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240

QY 241 AGCAGGACATAAGCGGGCTCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCATCTCTC 300

DB 241 AGCAGGACATAAGCGGGCTCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCATCTCTC 300

QY 301 TTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGCTGATTCATACCCACTCTAAAG 360  
Db 301 TTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGCTGATTCATACCCACTCTAAAG 360  
QY 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGAGAGTTAAATACATCAGAGA 420  
Db 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGAGAGTTAAATACATCAGAGA 420  
QY 421 TGATCAAGGAATTAAGGAATGCTTACCTCAGGAGGCTTATACAGATACGATATGTTAG 480  
Db 421 TGATCAAGGAATTAAGGAATGCTTACCTCAGGAGGCTTATACAGATACGATATGTTAG 480  
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGAGTGGCTCATG 540  
Db 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGAGTGGCTCATG 540  
QY 541 CCATGAATGAGTACCCAGACTCTGTCGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600  
Db 541 CCATGAATGAGTACCCAGACTCTGTCGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600  
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660  
Db 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660  
QY 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720  
Db 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720  
QY 721 GAATTGTGTAAGCAAGTGGTGCCTTAAGCAATCCAAACATTAACCAAACTCAATATG 780  
Db 721 GAATTGTGTAAGCAAGTGGTGCCTTAAGCAATCCAAACATTAACCAAACTCAATATG 780  
QY 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 840  
Db 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 840

RESULT 3

ID ABK47966 standard; cDNA; 841 BP.  
XX AC ABK47966;  
XX DT 02-JUL-2002 (first entry)  
XX DE Marine cDNA encoding G-CSF-inducible antibody binding protein, MMR19.  
XX KW Mouse; granulocyte-colony stimulating factor; G-CSF; ss; gene; MMR19;  
XX KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;  
XX KW infection.  
XX OS Mus sp.  
XX FH Key Location/Qualifiers  
XX CDS 6..731  
XX FT /\*tag= a  
XX FT /product= "G-CSF-inducible antibody binding protein"  
XX PN W0200226978-A1.  
XX  
XX PD 04-APR-2002.  
XX PF 27-SEP-2001; 2001WO-JP008446.  
XX PR 27-SEP-2000; 2000JP-00294191.  
XX PA (NIBS) JAPAN TOBACCO INC.  
XX PI Sha S, Mukai H, Aoki Y, Nishi Y;  
XX PI WPI; 2002-340016/37.  
XX DR P-PSDB; AAU71177.  
DR

XX Gene encoding protein binding to antibody having granulocyte-colony  
PT stimulating factor (G-CSF) inducing activity, useful for screening  
XX potential drugs treating G-CSF associated diseases.  
PS Claim 2; Page 91-93; 103pp; Japanese.  
XX  
CC The invention relates to a mouse or human gene (MMR19) encoding a protein  
CC which binds to antibodies or their fragments which induce granulocyte-  
CC colony stimulating factor (G-CSF) secretion. The genes and proteins of  
CC the invention are used in diagnosis, treatment and prevention of diseases  
CC associated with G-CSF, including infections and neutrophil deficiency  
CC disease. This sequence represents a cDNA encoding a mouse G-CSF-inducible  
CC antibody binding protein, MMR19  
XX  
SQ Sequence 841 BP; 274 A; 174 C; 203 G; 190 T; 0 U; 0 Other;

Query Match 100.0%; Score 840; DB 6; Length 841;  
Best Local Similarity 100.0%; Pred. No. 2e-183;  
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 60  
Db 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 60  
QY 61 ACAAGAGACACCCCGGATTCCTGATCCCAAGAACTTTGCAAAACAGTTTACCATCTGGGCT 120  
Db 61 ACAAGAGACACCCCGGATTCCTGATCCCAAGAACTTTGCAAAACAGTTTACCATCTGGGCT 120  
QY 121 GGGTCACTGGCTCGAGGGGGAACAGCTTCAAGCATGGCAATGAATCAATCTCATGCTC 180  
Db 121 GGGTCACTGGCTCGAGGGGGAACAGCTTCAAGCATGGCAATGAATCAATCTCATGCTC 180  
QY 181 CCTCAGGCGCTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240  
Db 181 CCTCAGGCGCTCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240  
QY 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCACCTCCTC 300  
Db 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCACCTCCTC 300  
QY 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGCTGATTCATACCCACTCTAAAG 360  
Db 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGCTGATTCATACCCACTCTAAAG 360  
QY 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGAGAGTTAAATACATCAGAGA 420  
Db 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGAGAGTTAAATACATCAGAGA 420  
QY 421 TGATCAAGGAATTAAGGAATGCTTACCTCAGGAGGCTTATACAGATACGATATGTTAG 480  
Db 421 TGATCAAGGAATTAAGGAATGCTTACCTCAGGAGGCTTATACAGATACGATATGTTAG 480  
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGAGTGGCTCATG 540  
Db 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGAGTGGCTCATG 540  
QY 541 CCATGAATGAGTACCCAGACTCTGTCGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600  
Db 541 CCATGAATGAGTACCCAGACTCTGTCGGGTTCTTGTCCGGGCTCATGGGTTGACGTGT 600  
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660  
Db 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660  
QY 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720  
Db 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720  
QY 721 GAATTGTGTAAGCAAGTGGTGCCTTAAGCAATCCAAACATTAACCAAACTCAATATG 780  
Db 721 GAATTGTGTAAGCAAGTGGTGCCTTAAGCAATCCAAACATTAACCAAACTCAATATG 780

Qy 781 CCTTAAATAAACTCAGCTGCTTTTAAAAA... 840  
 Db 781 CCTTAAATAAACTCAGCTGCTTTTAAAAA... 840

## RESULT 4

AAH75155

ID AAH75155 standard; cDNA; 1183 BP.

XX AAH75155;

XX AC

XX DT

XX 13-NOV-2001 (first entry)

XX Nucleotide sequence of a human enzyme.

XX Human; enzyme; cancer; neurological disorder; epilepsy; stroke;

XX Alzheimer's disease; Pick's disease; Huntington's disease; dementia;

XX multiple sclerosis; Parkinson's disease; amyotrophic lateral sclerosis;

XX meningitis; schizophrenia; neuroleptic disorder; allergy;

XX Addison's disease; autoimmune disease; anemia; asthma; Crohn's disease;

XX adult respiratory distress syndrome; atopic dermatitis; psoriasis;

XX diabetes mellitus; osteoporosis; pancreatitis; rheumatoid arthritis;

XX infection; genetic disorder; muscular dystrophy; Gaucher's disease;

XX Huntington's chorea; sickle cell anemia; thalassemia; atherosclerosis;

XX Von Willebrand's disease; Wilms' tumour; cell proliferative disorder;

XX leukemia; hepatitis; cirrhosis; arteriosclerosis; gene therapy; ss.

XX Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC arteriosclerosis). The polynucleotide is also useful in somatic or  
 CC germline gene therapy

XX SQ Sequence 1183 BP; 374 A; 213 C; 262 G; 334 T; 0 U; 0 Other;

Query Match 71.3%; Score 598.6; DB 5; Length 1183;

Best Local Similarity 84.9%; Pred. No. 6.9e-128;

Matches 694; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

Qy 4 CCATGCTGGCTGTCAAGCTC---AAGGAGACTGTTCTCGCGCCGCTGTGGCGCAGG 60

Db 92 CCATGCTGGCTGTGATGCTCGGGAGGAGACTGTTCTCCCGGAGATGGCGCGCAGG 151

Qy 61 ACAAGGAGACCCCGGATTCCTGATCCAGAACCTTTCGCAACAGTTTACCATCTGGCT 120

Db 152 ACAAGGAGCATCCAGATACCTGATCCAGAACCTTTCGCAACAGTTTACCATCTGGCT 211

Qy 121 GGGTCATGTCACCTGGAGGGGATCAGCTTGAAGCATGGCAATCAATCTACATTGCTC 180

Db 212 GGGTCATGTCACCTGGAGGGGATCAGCTTGAAGCATGGCAATCAATCTACATTGCTC 271

Qy 181 CCTCAGGCGTGCMAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTATG 240

Db 272 CTTTCAGGAGTGCAAAAGGAGCAATTCAGCTGAGGACATGTTTGTGTGATATAATG 331

Qy 241 AGCAGGACATTAAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCATCCTC 300

Db 332 AAAAGGACATAAGTGGACCTTCGCCATCGAAGAGAGCTAAAAAGGCGAGTGTACTCCTC 391

Qy 301 TTTTCATGATGCTTATACATGAGAGGAGCTGGCGGAGTGATTCATACCCACTCTAAAG 360

Db 392 TTTTCATGATGCTTATACATGAGAGGAGCTGGCGGAGTGATTCATACCCACTCTAAAG 451

Qy 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTAAAAATTCACATCAAGAGA 420

Db 452 CTGCTGTGATGGCCACCCCTTCTGTTCCAGACAGGAGTAAAAATTCACATCAAGAGA 511

Qy 421 TGATCAAGGATTAAGGAAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480

Db 512 TGATCAAGGATTAAGGAAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 571

Qy 481 TGGTACCTTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAGAGGATGGCTCATG 540

Db 572 TGGTACCTTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAGAGGATGGCTCATG 631

Qy 541 CCATGAATGATGATCCAGACCTCTGTCGGGTTCTTTCGGCGGCTCATGGGTTACGTTG 600

Db 632 CAATGAATGATGATCCAGACCTCTGTCGGGTTCTTTCGGCGGCTCATGGGTTACGTTG 691

Qy 601 GGGGAGAACATGGGAGAACCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 660

Db 692 GGGGAGAACATGGGAGAACCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 751

Qy 661 TTGCTGCTCCATGAAGAGATGGGACTTCGATCCAAACACAGCTCCCACTGGAGAAAAATG 720

Db 752 TTGCTGCTCCATGAAGAGATGGGACTTCGATCCCAACACAGCTCCCACTGGAGAAAAATG 811

Qy 721 GAATGTGTAAGCCCAAGTGGATGCTTAAGCATCTCCAAACAATTAAC-AAACTCAATTAT 779

Db 812 GAATGTGTAAGCCCAAGTGGATGCTTAAGCATCTCCAAACAATTAAC-AAACTCAATTAT 871

Qy 780 GCCTTAAATAAACTCAGCTGCTTTTAAAAA... 816

Db 872 TATTAAATGAAGCTATTTTTTAAATGAATGAAA 908

## RESULT 5

AAC61150

ID AAC61150 standard; DNA; 924 BP.

XX AAC61150;

XX AC

XX DT

XX 07-FEB-2001 (first entry)

WPI; 2001-550184/61.  
 P-PSDB; AAG67127.

Novel human enzyme molecule useful for treating and preventing, e.g.,  
 cancer, genetic disorders, neurological disorders, autoimmune and  
 inflammatory disorders.

Claim 5; Page 140; 154pp; English.

The present sequence encodes a human enzyme. The enzyme polynucleotide  
 and polypeptide are useful for diagnosis, treatment and prevention of  
 cancers, neurological disorders (e.g. epilepsy, stroke, Alzheimer's  
 disease, Pick's disease, Huntington's disease, dementia, multiple  
 sclerosis, Parkinson's disease, amyotrophic lateral sclerosis, bacterial  
 and viral meningitis, schizophrenia disorders and neuroskeletal  
 disorders), autoimmune/inflammatory disorders (e.g. allergies, Addison's  
 disease, autoimmune diseases, adult respiratory distress syndrome,  
 anemia, asthma, Crohn's disease, atopic dermatitis, diabetes mellitus,  
 osteoporosis, pancreatitis, psoriasis, rheumatoid arthritis, and viral,  
 bacterial, fungal, parasitic, protozoal and helminthic infections),  
 genetic disorder (e.g. Duchenne and Becker muscular dystrophy, Gaucher's  
 disease, Huntington's chorea, sickle cell anemia, thalassemia, Von  
 Willebrand's disease and Wilms' tumour), and cell proliferative disorder  
 (e.g. atherosclerosis, leukemia, hepatitis, cirrhosis, and



XX DE DNA encoding antigen recognised by Ab capable of inducing G-CSF activity.  
XX KW Antigenic protein; antibody; granulocyte colony stimulating factor;  
XX KW G-CSF; cancer therapy; bone marrow suppression; human; ds.  
OS Homo sapiens.  
PN WO200060075-A1.  
XX 12-OCT-2000.  
XX 31-MAR-2000; 2000WO-JP002080.  
XX 01-APR-1999; 99JP-00095092.  
XX (NIBS) JAPAN TOBACCO INC.  
XX Sha S, Aoki Y, Nishi Y;  
PI WPI; 2001-024452/03.  
DR P-PSDB; AAY85636.  
XX  
PT Gene encoding an antigen recognizing an antibody which induces  
PT granulocyte colony stimulating factor (G-CSF) expression for gene therapy  
PT and treatment of G-CSF associated disorders e.g. the side effects of  
PT cancer therapy.  
XX  
XX Claim 4; Page 50-52; 58pp; Japanese.  
XX  
XX The present invention relates to a gene encoding an antigenic protein  
CC recognised by an antibody or its fragments which can induce the  
CC production of granulocyte colony stimulating factor (G-CSF). Also  
CC included in the invention are partial sequences of the gene, antibodies  
CC recognising all or part of the antigenic protein, expression vectors  
CC containing the gene and host cells transformed by the vector. The gene is  
CC used for gene therapy, and compounds identified by screening using the  
CC gene sequence are used for treatment and prevention of disorders  
CC associated with G-CSF expression such as the side effects of cancer  
CC therapy (including bone marrow suppression). The present sequence  
CC represents the human gene of the invention  
XX  
SQ Sequence 924 BP; 291 A; 166 C; 225 G; 242 T; 0 U; 0 Other;  
Query Match 70.9%; Score 595.4; DB 5; Length 924;  
Best Local Similarity 84.7%; Pred. No. 3.5e-127;  
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;  
4 CCATGCTGGCTGCTCAAGCT---CAAGGAGACTGTTGCTCGCGCGCTGTGGCGCGCAGG 60  
44 CCATGCTGGCTGCTGATGCTTGGGAGGAGACTGTTGTTCCCGGAGATGGCGCGCAGG 103  
61 ACAAGAGACACCCCGATTCTGATCCCAAGACTTTCGCAACAGTTTACCATCTGGCT 120  
104 ACAAGGAGCATCCAAGATACCTGATCCAGAACTTTTCGCAACAGTTTACCATCTAGGCT 163  
121 GCGTCACTGCACTGGAGGGGAAATGATGTAAGCATGCAATGAATCTACATTGCTC 180  
164 GGGTCACTGGGACTGGAGGAGAAATGATGTAAGCATGCGATGGAATCTACATTGCTC 223  
181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCGAGAGACATGTTTGTGTGACATTATG 240  
224 CTTTCAGGAGTGCATAAGAGACGAAATTCAGCTGAGAGACATGTTTGTGATGATATAATG 283  
241 AGCAGGACATAAGCGGCGCTTCAGCATCTTAAGAGCTGAAAAAGCAGTGCATCTC 300  
284 AAAAGGACATAAGTGGACCTTCGCTCATCGAAGAGCTTAAAAAAGCAGTGTACTCCTC 343  
301 TTTTCATCATGCTTATACATGAGAGGAGCTGGCGAGTGATTCATACCCACTCTAAAG 360  
344 TTTTCATGATGCTTACATGAGAGAGAGAGAGTGAGTGATTCATACCCACTCTAAAG 403  
361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAAAATTACATCAAGAGA 420

Db 404 CTGCTGTGATGCCACCTTCTCTTTCCAGGACGGAGTTTAAAAATTACATCAAGAGA 463  
QY 421 TCATCAAAAGGAAATAGGAAATGTACTCTCAGGAGGCTATTACAGATACGATGATGTTAG 480  
Db 464 TGATTAAGGAAATAGGAAATGTACTCTCAGGAGGCTATTATAGATATGATGTTAG 523  
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGATCTCAAGAGAGATGGCTCATG 540  
Db 524 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGATCTCAAGAGAGATGGCTCATG 583  
QY 541 CCATGAATGAGTACCCAGACTCTCTGTCGGGTTCTTGTCCGGGCTCATGGGTTGAGCTG 600  
Db 584 CAATGAATGAATACCCAGACTCTCTGTCAGTACTGCTCAGACGTCATGAGTATATGTT 643  
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACTCTGTTTGACA 660  
Db 644 GGGGGGAAACATGGGAGAAAGGCAAAACCAATGTGTGAGTGTATTGACTACTCTGTTTGATA 703  
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAAAATG 720  
Db 704 TTGCGGTATCAATGAAGAAAGTAGGACTTGTATCTTCCAGCTCCAGTTGGAGAAAAATG 763  
QY 721 GAATGTGTACCCAGTGTGATGCTTACGATCTCCAAACAATAAAC-AACTCAATTAT 779  
Db 764 GAATGTGTACCCAGTGTGATGCTTAAATATATATACAGAGATAAAGCTAAACGTAATTAT 823  
QY 780 GCCTTAATAATAAATCAAGCTGCTTTTAAAAAATAAAAA 816  
Db 824 TATTTAATGAAGCTATTTTTTTTAAATGATGAAA 860  
RESULT 6  
ABK47967  
ID ABK47967 standard; cDNA; 1136 BP.  
XX AC ABK47967;  
XX DT 02-JUL-2002 (first entry)  
XX DE Human cDNA encoding G-CSF-inducible antibody binding protein, MMR19.  
XX KW Human; granulocyte-colony stimulating factor; G-CSF; ss; gene; MMR19;  
XX KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 46..774  
XX FT /\*tag= a  
XX FT /product= "G-CSF-inducible antibody binding protein"  
XX PN WO200226978-A1.  
XX PD 04-APR-2002.  
XX PF 27-SEP-2001; 2001WO-JP008446.  
XX PR 27-SEP-2000; 2000JP-00294191.  
XX PA (NIBS) JAPAN TOBACCO INC.  
XX PI Sha S, Mukai H, Aoki Y, Nishi Y;  
XX DR WPI; 2002-340016/37.  
XX DR P-PSDB; AAY77178.  
XX PT Gene encoding protein binding to antibody having granulocyte-colony  
XX PT stimulating factor (G-CSF) inducing activity, useful for screening  
XX PT potential drugs treating G-CSF associated diseases.  
XX PS Claim 4; Page 94-96; 103pp; Japanese.

XX The invention relates to a mouse or human gene (MMR19) encoding a protein  
CC which binds to antibodies or their fragments which induce granulocyte-  
CC colony stimulating factor (G-CSF) secretion. The genes and proteins of  
CC the invention are used in diagnosis, treatment and prevention of diseases  
CC associated with G-CSF, including infections and neutrophil deficiency  
CC disease. This sequence represents a cDNA encoding a human G-CSF-inducible  
CC antibody binding protein, MMR19  
XX  
SQ Sequence 1136 BP; 371 A; 188 G; 248 G; 329 T; 0 U; 0 Other;  
Query Match 70.9%; Score 595.4; DB 6; Length 1136;  
Best Local Similarity 84.7%; Pred. No. 3.7e-127;  
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;  
QY 4 CCATGCTGGCTGTCAAGCT---CAAGGAGACTGTGCTCGCGCGCTGTGCGGCGAGG 60  
Db |||||  
QY 44 CCATGCTGGCTGTGATGCTCGGAGGGAGACTGTGTTCCCGGAGATCGCGCGCGAGG 103  
Db |||||  
QY 61 ACAAGGAGCACCCTGATTCCTGATCCCAAGACTTTGCAACAGATTTTACCATCTGGCT 120  
Db |||||  
QY 104 ACAAGGAGCATCCAGATACCTGATCCCAAGACTTTGCAACAGATTTTACCATTTAGCT 163  
Db |||||  
QY 121 GGGTCACTGGCACTGAGGGGAAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180  
Db |||||  
QY 164 GGGTCACTGGCACTGAGGGGAAATTAGCTTGAAGCATGGCAATGAAATCTACATGCTC 223  
Db |||||  
QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240  
Db |||||  
QY 224 CTTTCAGGAGTGCAAAAGGAGCAATTTTTCAGCCAGAGCATGTTGTTTATGATATAATG 283  
Db |||||  
QY 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGATGCTCCTC 300  
Db |||||  
QY 284 AAAGGACATAGTGAGCTTCGCCATCGAAGAGCTTAAAAAGCCAGTGTACTCCTC 343  
Db |||||  
QY 301 TTTTCATGAATGTTATACATGAGAGAGCTGGCGAGTGATTCATCCACTCTAAAG 360  
Db |||||  
QY 344 TTTTCATGAATGTTATACATGAGAGAGCTGGCGAGTGATTCATCCACTCTAAAG 403  
Db |||||  
QY 361 CTGCTGTGAGTACCTCTGTTTCCAGGACAGAGTTTAAATTTACACATCAAGAGA 420  
Db |||||  
QY 404 CTGCTGTGAGTACCTCTGTTTCCAGGACAGAGTTTAAATTTACACATCAAGAGA 463  
Db |||||  
QY 421 TGATCAAGAAATGAAGAAATGATACCTCAGAGGCTTATACAGATACGATGATGTTAG 480  
Db |||||  
QY 464 TGNATAAGAAATGAAGAAATGATCTCCGAGGGTATTATAGATGATGATGTTAG 523  
Db |||||  
QY 481 TGGTACCTATTATTGAGAACTCTGAGAGAGAGATCTCAAAGAAAGATGGCTCATG 540  
Db |||||  
QY 524 TGGTACCTATTATTGAGAACTACCTGAGAGAGAGACCTCAAAGATAGAAATGGCTCATG 583  
Db |||||  
QY 541 CCATGAATGATGATGAGTCTGCTGCGGTTCTTCTCGGCGTCAATGGGTTGATGTTG 600  
Db |||||  
QY 584 CAATGAATGATGATGAGTCTGCTGCGGTTCTTCTCGGCGTCAATGGGTTGATGTTG 643  
Db |||||  
QY 601 GGGGAGAAACATCGGAGAGAGCAAAACCATGTTGATGATGATGATGATGATGATGATG 660  
Db |||||  
QY 644 GGGGAGAAACATCGGAGAGAGCAAAACCATGTTGATGATGATGATGATGATGATGATG 703  
Db |||||  
QY 661 TTGCTGCTCTCCATGAGAGAGATGGGACTCGATCCAAACAGCTCCCGATGGGAGAAATG 720  
Db |||||  
QY 704 TTGCTGCTCTCCATGAGAGAGATGGGACTCGATCCAAACAGCTCCCGATGGGAGAAATG 763  
Db |||||  
QY 721 GAATTTGTTGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779  
Db |||||  
QY 764 GAATTTGTTGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 823  
Db |||||  
QY 780 GCCTTAATAAATCTAGCTGCTTTTAAAAAATAA 816  
Db |||||  
QY 824 TATTTAATGAAGCTATTTTAAATGAATGAATGA 860  
Db |||||

AAAX05748/c  
ID AAAX05748 standard; cDNA; 1190 BP.  
XX  
AC AAAX05748;  
XX  
DT 28-APR-1999 (first entry)  
XX  
DE Nucleotide sequence of human HFI2G53.  
XX  
KW HFI2G53; human; inflammatory disease; infection; HIV-1; HIV-2; cancer;  
KW HIV-associated cachexia; immunodeficiency disorder; septic shock; pain;  
KW Parkinson's disease; cardiovascular disease; psychotic; neurological;  
KW Huntington's disease; Gilles de la Tourette's syndrome; gene mapping;  
KW mental retardation; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 154..879  
FT /\*tag= a  
FT /note= "the specification indicates that this region  
FT encodes the human HFI2G53 protein (AAW94762); the  
FT translated product of this sequence does not match the  
FT protein sequence AAW94762"  
XX  
PN EP892050-A2.  
XX  
PD 20-JAN-1999.  
XX  
PF 17-FEB-1998; 98EP-00301168.  
XX  
PR 08-JUL-1997; 97US-0051937P.  
PR 17-OCT-1997; 97US-00953494.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Demarini D;  
XX  
DR WPI; 1999-083567/08.  
DR P-PSDE; AAW94762.  
XX  
PT New HFI2G53 polypeptide and polynucleotide - useful as diagnostic  
PT reagents and for prevention and treatment of inflammatory diseases,  
PT cancer and Parkinson's disease.  
XX  
PS Claim 2; Page 6-7; 22pp; English.  
XX  
CC This represents the nucleotide sequence of human HFI2G53. Host cells  
CC containing an expression system comprising the HFI2G53 nucleic acid are  
CC used for the recombinant production of the protein. HFI2G53 polypeptides  
CC and polynucleotides are useful for diagnosing diseases related to over or  
CC underexpression of HFI2G53 protein. The HFI2G53 polypeptides can be used  
CC to screen for agonists and antagonists which can be used in treatment to  
CC activate or inhibit HFI2G53 activity. Gene therapy may also be used to  
CC affect endogenous polypeptide production, using HFI2G53 polynucleotides  
CC and retroviral vectors. HFI2G53 antibodies are useful for inducing an  
CC immune response to immunise and prevent diseases, and for isolating  
CC HFI2G53 clones or purifying the polypeptide by affinity chromatography.  
CC HFI2G53 polypeptides can be administered directly or as a vaccine to  
CC inoculate against disease. Diseases prevented, diagnosed or treated  
CC include inflammatory diseases such as Adult Respiratory Disease Syndrome,  
CC rheumatoid arthritis, osteoarthritis, inflammatory bowel disease, asthma,  
CC psoriasis, dermatitis, allergies; infections including bacterial, fungal,  
CC protozoan and viral, particularly HIV-1 and -2; HIV-associated cachexia  
CC and other immunodeficiency disorders; septic shock; injury; pain; cancers  
CC including testicular cancer; anorexia; bulimia; Parkinson's disease;  
CC cardiovascular disease including restenosis, atherosclerosis, acute heart  
CC failure, myocardial infarction, hypertension, hypertension; urinary  
CC retention; angina pectoris; ulcers; benign prostatic hypertrophy; and  
CC psychotic and neurological disorders (anxiety, schizophrenia, delirium,  
CC manic depression, dementia, severe mental retardation) and dyskinesias,  
CC such as Huntington's disease or Gilles de la Tourette's syndrome. The  
CC HFI2G53 polypeptide is also useful for mapping the gene to a chromosome,  
CC

CC allowing gene inheritance to be studied through linkage analysis

XX Sequence 1190 BP; 317 A; 287 C; 240 G; 346 T; 0 U; 0 Other;

Query Match 70.9%; Score 595.4; DB 2; Length 1190;  
Best Local Similarity 84.6%; Pred. No. 3.8e-127;  
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

```
QY 4 CCATGCTGGCTGCTCAAGCT---CAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 995 CCATGCTGGCTGCTGCTGTTGGAGGAGACTGTTGTTCCGGAGATGCGCGCGCAGG 936
QY 61 ACAAGAGACACCCCGATTCCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 120
DB 935 ACAAGAGACATCCAGATACCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 876
QY 121 GGCTCACTGACCTGGAGGGGATCAGCTGAGCATGCAATGAATCTACATGCTC 180
DB 875 GGCTCACTGAGCTGGAGGAGGATGAGCTTGAAGCATGCGGATGAATCTACATGCTC 816
QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240
DB 815 CTTCAGAGGTGCAAAAGGAGCGAATTCAGCTGAGACATGTTGTTTATGATATAATG 756
QY 241 AGCAGGACATAAGCGGCCCTCCAGCATCTAAGAGCTGAAAAAGCCAGCTGCTCCTC 300
DB 755 AAAAGGACATAAGTGGACCTTCGCCATCGAAGAGCTTAAAAAGGCGAGTACTCTCCTC 696
QY 301 TTTTCATGATGCTTATACATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
DB 695 TTTTCATGATGCTTATACATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 636
QY 361 CTGCTGTGATGGCTTACCTCTCTGTTCCAGACAGGAGTTTAAATTTACACATCAAGAGA 420
DB 635 CTGCTGTGATGGCCACCTCTCTCTTCCAGGACGGAGTTTAAATTTACACATCAAGAGA 576
QY 421 TGATCAAGGAATTAAGAAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
DB 575 TGATAAAGGAATTAAGAAATGTACTTCCGAGGCTATTATAGATATGATGATGTTAG 516
QY 481 TGGTACCTATTATGAGAACTCTCTGAGAGAGGATCTCAAGAGAGGATGCTCATG 540
DB 515 TGGTACCTATTATGAGAACTCTCTGAGAGAGGATCTCAAGAGAGGATGCTCATG 456
QY 541 CCATGATGATGATCCAGACTCTGTCGGGTTCTGTCCGCGCTCATGGGCTGTACGCTGT 600
DB 455 CAATGATGATATCCAGACTCTGTCGAGTACTGTCAGAGCTCATGGATATATGTTGT 396
QY 601 GGGAGAGAAACATGGAGAGAAAGCAAAACCATGTGTGAGTGTATGACTCTGTTTGACA 660
DB 395 GGGGGAGAAACATGGAGAGAGGCAAAACCATGTGTGAGTGTATGACTATTTATTGATA 336
QY 661 TTGCTGTCTCCATGAGAGATGGAGCTCGATCCATCAACAGACTCCAGTTCGAGAAATG 720
DB 335 TTGCGGTATCAATGAAGAAAGTAGGACTTGATCTTCCAGCTCCAGTTCGAGAAATG 276
QY 721 GAATGTGTAGCAAGTGGATGCTTGAAGATCTCCAAACATAAAGC-AAAACCTCAATAT 779
DB 275 GAATGTGTAGCAAGTGGATGCTTGAAGATCTTATATATACAGAGTAAAGCTTAATAT 216
QY 780 GCCTTAATAAATCTCAGCTCTCTTTTAAAAAAGAAAA 816
DB 215 TATTAAATGAAGCTATTTTTTTTAAATGAATTGAAA 179
```

RESULT 8  
ABK12566  
ID ABK12566 standard; cDNA, 1226 BP.  
XX ABK12566;  
XX  
XX  
XX 18-JUN-2002 (first entry)  
XX

DNA encoding cell differentiation stimulator associated protein #1.

Cartilage cell differentiation stimulator; osteopathic;  
Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;  
Membrane bound type transferrin-like protein; Mtf; cartilage disorder;  
bone metabolism disease; cell differentiation; cell growth;  
extracellular matrix related disease; gene; ss; human.

Homo sapiens.  
Location/Qualifiers  
Key 78, 806  
CDS  
/tag= a  
/product= "Cell differentiation stimulator associated protein #1"

JP2002C20311-A.  
23-JAN-2002.  
07-JUL-2000; 2000JP-00206566.  
07-JUL-2000; 2000JP-00206566.  
(KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.

WPI; 2002-287405/33.  
P-PSDB; AAU78360.

A cartilage cell differentiation stimulator useful in the diagnosis of  
biophylaxis, cell differentiation, cell growth and construction of  
extracellular matrix related diseases.

Claim 5; Page 8; 17pp; Japanese.

The invention describes a cartilage cell differentiation stimulator  
(containing a membrane-bound transferrin-like protein (Mtf-BP) and a  
membrane bound type transferrin-like protein (Mtf)) and an animal-derived  
concanavalin-like drug. The cartilage differentiation stimulator can be  
used in diagnosis, prevention and treatment of cartilage and bone  
metabolism diseases. They can also be used for diagnosing biophylaxis,  
cell differentiation, cell growth and construction of extracellular  
matrix related diseases. Mtf-BP strongly stimulates differentiation of  
cartilage cells and exhibits similar action mechanism with that of plant  
derived ConA. This sequence represents a cartilage cell differentiation  
stimulator associated polypeptide described in the invention

Sequence 1226 BP; 382 A; 212 C; 274 G; 358 T; 0 U; 0 Other;

Query Match 70.7%; Score 593.8; DB 6; Length 1226;  
Best Local Similarity 84.6%; Pred. No. 8.9e-127;  
Matches 691; Conservative 0; Mismatches 122; Indels 4; Gaps 2;

```
QY 4 CCATGCTGGCTGCTCAAGCT---CAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 76 CCATGCTGGCTGCTGATGCTGGGAGGAGACTGTTGTTCCGGAGATGCGCGCGCAGG 135
QY 61 ACAAGAGACACCCCGATTCCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 120
DB 136 ACAAGAGACATCCAGATACCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 195
QY 121 GGCTCACTGACCTGGAGGGGATCAGCTTGAAGCATGCGAATGAATCTACATGCTC 180
DB 196 GGCTCACTGCGACTGGAGGAGGAAATAGCTTTGAAGCATGCGGATGAATCTACATGCTC 255
QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240
DB 256 CTTCAGGAGTGCAAAAGGAGGAAATTCAGCTGAGACATGTTGTTTGTGATATAATG 315
QY 241 AGCAGGACATAAGCGGCCCTCCAGCATCTAAGAGCTGAAAAAGGCGAGTACTCTCCTC 300
DB 316 AAAAGGACATAAGTGGAGCTTCGCCATCGAAGAGCTTAAAAAGGCGAGTACTCTCCTC 375
```



Db 122 CCATGAAGAGATGGGCTGGATCCCAACGCGAGTTCCTCCAGTTGGAGAACATGGAAATCGTAT 63  
QY 730 AAGCCAGTGGATGCTTACGATCTCCACAAATAAACAACACTCAATTATGCTTAAATA 789  
Db 62 AAGCCAGTGGAGCGCTTAAGATCTCCAAATAAACAACACTCAATTATGCTTAAATA 3  
QY 790 AA 791  
Db 2 AA 1  
RESULT 10  
ID ABX63252/c  
XX ABX63252 standard; cDNA; 3764 BP.  
AC ABX63252;  
XX 25-FEB-2003 (first entry)  
DT Human cDNA #252 differentially expressed in activated vascular tissue.  
DE Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hypertension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.  
XX Homo sapiens.  
OS  
XX US2002137081-A1.  
PN  
XX 26-SEP-2002.  
PD  
XX 08-JAN-2002; 2002US-00044090.  
XX 28-JUL-2000; 2000US-0222469P.  
PR 08-JAN-2001; 2001US-0260483P.  
PR (BAND/) BANDMAN O.  
PA Bandman O;  
PI WPI; 2003-110597/10.  
DR  
XX Combination for diagnosing, staging, treating, or monitoring the  
PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
PT comprises several cDNAs that are differentially expressed in activated  
PT vascular tissue.  
XX Claim 1; Page; 18pp; English.  
XX This invention relates to a combination comprising several cDNAs that are  
CC differentially expressed in activated vascular tissue. The invention also  
CC discloses a high throughput method for detecting differentially expressed  
CC cDNAs in a sample. The cDNAs of the invention may have  
CC antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;  
CC gynaecological; vasotropic and cerebroprotective activities and may be  
CC used in gene therapy. The cDNAs of the invention may be used in a high-  
CC throughput methods for detecting differential expression of one or more  
CC cDNAs in a sample, or screening several molecules or compounds to  
CC identify a molecule or compound that specifically binds a cDNA of the  
CC invention. A protein encoded by the cDNA may be used to screen several  
CC molecules or compounds to identify a ligand that specifically binds to  
CC the protein, or to produce or purify an antibody to the protein that can  
CC be used to detect a protein in a sample or purify a natural or  
CC recombinant protein from a sample. The nucleotides may be useful for  
CC diagnosing, staging, treating, or monitoring the progression of treatment  
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion  
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
CC genetic or gene expression analysis of several new nucleic acid  
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for

CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
CC associated with abnormalities in the expression, amount or distribution  
CC of the protein. The present sequence represents a cDNA of the invention  
CC that is differentially expressed in activated vascular tissue. Note: The  
CC sequence data for this patent did not form part of the specification, but  
CC was obtained in electronic format directly from USPTO at  
CC http://seqdata.uspto.gov/sequence.html?docID=20020137081  
XX  
SQ Sequence 3764 BP; 1175 A; 758 C; 777 G; 1053 T; 0 U; 1 Other;  
Query Match 47.6%; Score 399.8; DB 7; Length 3764;  
Best Local Similarity 79.9%; Pred. No. 4.6e-82;  
Matches 671; Conservative 0; Mismatches 143; Indels 26; Gaps 16;  
QY 4 CCATGTCTGGCTGTCAAGCTCAAGGAG-----ACTGTGTCTCGCGCGCTGTGCGCGCAG 59  
Db 1182 CCATGTCTGGCTGTGATGCTCGGAGGGAGACTGGTGTTCGCGAGATGCGCGCGCAG 1123  
QY 60 GACAAGGAGACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGC 119  
Db 1122 GACAAGGAGATCCAGATACCTGATCCAGAACTTTGCAACAGTTTACCATTTAGGC 1063  
QY 120 TGGGTCCAC--TGSCACTGGAGGGGGAATCAGCTTTGAAGCATGGC-AATGAAATCTCAT 176  
Db 1062 TGTGTCACTGTGTCACCTGGAGGAGGAATTAGCTTTGAAGCATGGGATTGAAATCTCAT 1003  
QY 177 GCTCCCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGT-GTGACAT 235  
Db 1002 GCTCCTTCAGGAGTGCAAAAGGAGCAATTCAGCTTGAAGCATGTTTGTATTGATAT 943  
QY 236 TAATGAGCGGACATAAGCGGG--CTTCAGCATCTTAAGAGCT-GAAAGGAGCGAGTGC 293  
Db 942 AATGAAAGGACATAAGTGGGACCTTCGCAATCGAAGAGCTTAAAGGAGCGAGTGT 883  
QY 294 ACTCCTC-TTTTCATCAATGCTTATACCATGAGAGGCTGGCGCAGTGATTCATACCCA 352  
Db 882 ACTCCTCTTTTTCATGATGCTTACCAATGAGAGGAGGCTGGCGAGTTCATACCCA 823  
QY 353 CTCCTAAAGCTGTGTGATGCTACCTCTCT-GTTCCAGGACAGGAGTTTAAATTTACAC 411  
Db 822 CTCCTAAAGCTGTGTGATGCTACCTCTCTCTTTTCCAGGACGCGAGTTTAAATTTACAC 763  
QY 412 ATCAAGAGATGATCAAGGAATAAGGAATGTACTCTCAGAGGCTATTTACAGATACATG 471  
Db 762 ATCAAGAGATGATAAAGGAATAAAGAAATGTACTTCCGAGGGGTATTTAGATATGATG 703  
QY 472 ATATGTAGTGGT-ACCTATTATTGAGACACTCTCTGA-AGAGAAGATCTCAAGAAAG 529  
Db 702 ATATGTAGTGGTGGGCTCCCTATTTAGATTTCCACCTGAGGAGAAAGACCTCAAGATTG 643  
QY 530 GA--TGGCTCATGCTCATGAATG--AGTACCAGACTCC---TGTGGGGTTCCTTGTCCGGC 582  
Db 642 GAATGGCTCATGCTCATGAATGGAATACCCAGACTCCCTGTNCCAGTACTGTGTAGAC 583  
QY 583 GT-CATGGGTGTACGTGTGGGAGAAACATGGGAGAAAG--CAAAAACCATGTGTGAGT 639  
Db 582 GTCCATGAGTATATGTGTGGGGGAAACCATGGGAGAGGGCCCAAAACCATGTGTGAGT 523  
QY 640 GTTATGACTACTGTTGTGACATTCCTGTCTCCATGAAGAGATGGGACTCGATCCCAAC 699  
Db 522 GTTATGACTATTTTATTTGATATTTGCGGTATCAATGAAGAAAGTAGGACTTGTATCTTAC 463  
QY 700 AGCTCCAGTGTGAGAAATGGAATTTGTGAAGCAAGTGAATGCGCTCAAGCATCTCCAAC 759  
Db 462 AGCTCCAGTGTGAGAAATGGAATTTGTCTAAGCCAAAGAAAGTCTAATTATATACAGA 403  
QY 760 AATFAAAC--AAACTCAATTATGCCCTTAATAAATACTAGCTGCTTTTAAAAAATAAAAA 817  
Db 402 GATAAAGTAAACGTTAAATTTATTTTAAATGAAGCTATTTTAAAAATGAAGATTGAAA 343  
RESULT 11  
AAZ16856

PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stachne-Crain B;  
XX WPI; 1999-494092/41.  
XX  
XX  
XX Novel human genes and their expression products which are differentially  
XX expressed in different cell types.  
XX  
XX Claim 1; Page 2048; 2479pp; English.  
XX  
XX The present invention describes a library of human polynucleotides  
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
XX a method of detecting differentially expressed genes correlated with the  
XX cancerous state of a mammalian cell, comprising detecting at least one  
XX differentially expressed gene product in a test sample from a cell  
XX suspected of being cancerous, where the gene product is encoded by one of  
XX the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
XX polynucleotides can be used as a source of primers and probes, which can  
XX be used for a variety of purpose, e.g. detection of expression levels,  
XX mapping, tissue typing or profiling, forensics, genetic analysis and  
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides  
XX can be used for raising antibodies for experimental, diagnostic and  
XX therapeutic purposes. The polynucleotides may also be used to construct  
XX arrays for diagnostics (which may be used to determine function of an  
XX encoded protein); and to detect differences in expression levels between  
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to  
XX identify a genetic predisposition or susceptibility to a disease such as  
XX cancer). The polynucleotides of the invention are especially used in the  
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,  
XX and lung cancer. The polynucleotides can also be used to screen for  
XX peptide analogues and antagonists  
XX  
XX Sequence 758 BP; 241 A; 122 C; 146 G; 233 T; 0 U; 16 Other;  
XX  
XX Query March 42.9%; Score 360; DB 2; Length 758;  
XX Best Local Similarity 80.6%; Pred. No. 4.5e-73;  
XX Matches 431; Conservative 0; Mismatches 103; Indels 1; Gaps 342  
XX 283 AAAGCAGTGCACCTCTTTTTCATGTAATGCATTATCATTACGAGAGAGCTGCGCAGTGA

RESULT 12	
AAZ17165	
ID	AAZ17165 standard; cDNA; 778 BP.
XX	
XX	
AC	AAZ17165;
XX	
XX	
DT	12-OCT-1999 (first entry)
XX	
XX	Human gene expression product cDNA sequence SEQ ID NO:4636.
XX	
KW	Human; gene; gene expression product; diagnosis; therapy; probe;
KW	detection; mapping; tissue typing; profiling; forensic; cancer;
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO938972-A2.
XX	
XX	05-AUG-1999.
PD	
XX	
PF	28-JAN-1999; 99WO-US001619.
XX	
PR	28-JAN-1998; 98US-0072910P.
PR	24-FEB-1998; 98US-0075954P.
PR	31-MAR-1998; 98US-0080114P.
PR	03-APR-1998; 98US-0080515P.
PR	03-APR-1998; 98US-0080866P.
PR	21-OCT-1998; 98US-0105234P.
PR	28-OCT-1998; 98US-0105877P.
XX	
FA	(CHIR ) CHIRON CORP.
PA	(HYSE-) HYSEQ INC.
XX	
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI	Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI	Leschkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX	

DR WPI; 1999-494092/41.  
 XX Novel human genes and their expression products which are differentially  
 PT expressed in different cell types.  
 PS Claim 1; Page 2200; 2479pp; English.  
 XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one of  
 CC the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists  
 XX Sequence 778 BP; 244 A; 129 C; 150 G; 241 T; 0 U; 14 Other;  
 SQ Query Match 42.1%; Score 354; DB 2; Length 778;  
 Best Local Similarity 81.1%; Pred. No. 1.1e-71;  
 Matches 423; Conservative 0; Mismatches 96; Indels 2; Gaps 1;  
 QY 298 CTCTTTTCATGATGCTTATACATGAGAGAGAGTGGCGGAGTTCATACACCTCTA 357  
 DB |||||  
 46 CTTTNTGAGGATCCCATCGATTCGAGAGAGAGGAGTGCAGTGCATTCATACACCTCTA 105  
 QY 358 AGCTGCTGTGATGGCTACCTCTGTTTCCAGACAGGAGTTTAAATTCACATCAAG 417  
 DB |||||  
 106 AAGTGTGCTGTGATGGCCACCCCTCTCTTTCCAGACGGGAGTTTAAATTCACATCAAG 165  
 QY 418 AGATGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGT 477  
 DB |||||  
 166 AGATGATCAAGGAATAGGAATGTACTTCGAGGGTATTATAGATATGATGATGT 225  
 QY 478 TAGTGGTACCTATTATTGAGACAACCTCCTGAGAGAGAGATCTCAAGAAAGGATGGCTC 537  
 DB |||||  
 226 TAGTGGTACCTATTATTGAGAAATACACCTCAGGAGAGAGACCTCAAGATAGATGGCTC 285  
 QY 538 ATGCCATGATGATGATCCAGACTCCTGCTGGTCTTGTCCGCGCTCATGGGTGTACG 597  
 DB |||||  
 286 ATGCAATGATGATGATATCCAGACTCCTGTGCTGATGCTGATGATGATGATGATG 345  
 QY 598 TGTGGGAGAGAAATGGAGAGAGAAACCAATGTGTGATGTGTATGATCTGTTG 657  
 DB |||||  
 346 TGTGGGAGAGAAATGGAGAGAGAGCAACCAATGTGTGATGTGTATGATCTTATTG 405  
 QY 658 ACATTGTGCTTCCATGAAGAAGATGGAGCTCGATCCAAACACAGCTCCAGTTGGAGAAA 717  
 DB |||||  
 406 ATATTGCGGTATCAATGAAGAAGATGAGACTTTGATCTTCACAGCTCCAGTTGGAGAAA 465  
 QY 718 ATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 775  
 DB |||||  
 466 ATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525  
 QY 776 TTATGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 816  
 DB |||||  
 526 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 566

RESULT 13

AAZ15675  
 ID AAZ15675 standard; cDNA; 851 BP.  
 XX AAZ15675;  
 AC AAZ15675;  
 XX 12-OCT-1999 (first entry)  
 DT Human gene expression product cDNA sequence SEQ ID NO:3144.  
 DE Human; gene; gene expression product; diagnosis; therapy; probe;  
 XX detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; es.  
 XX Homo sapiens.  
 OS WO9938972-A2.  
 XX 05-AUG-1999.  
 PD 28-JAN-1999; 99WO-US001619.  
 PF 28-JAN-1999; 98US-0072910P.  
 PR 24-FEB-1998; 98US-0075954P.  
 PR 31-MAR-1998; 98US-0080114P.  
 PR 03-APR-1998; 98US-0080515P.  
 PR 03-APR-1998; 98US-0080666P.  
 PR 21-OCT-1998; 98US-0105234P.  
 PR 28-OCT-1998; 98US-0105877P.  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Ckvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;  
 XX WPI; 1999-494092/41.  
 DR Novel human genes and their expression products which are differentially  
 PT expressed in different cell types.  
 PS Claim 1; Page 1511; 2479pp; English.  
 XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one of  
 CC the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists  
 XX Sequence 851 BP; 266 A; 143 C; 162 G; 243 T; 0 U; 37 Other;  
 SQ Query Match 41.5%; Score 348.2; DB 2; Length 851;  
 Best Local Similarity 81.1%; Pred. No. 2.4e-70;  
 Matches 415; Conservative 0; Mismatches 96; Indels 1; Gaps 1;  
 QY 308 GATGCTTATACCATGAGAGAGTGGCGGAGTTCATACACCTCAAGCTGTGT 367







Search completed: June 1, 2004, 22:28:39  
Job time : 556 secs

PA (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 3484; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 796 BP; 194 A; 214 C; 220 G; 168 T; 0 U; 0 Other;  
Query Match 30.5%; Score 256; DB 4; Length 796;  
Best Local Similarity 62.5%; Pred.No. 3.7e-49;  
Matches 400; Conservative 0; Mismatches 240; Indels 0; Gaps 0;  
QY 65 GGAGCACCCCGGATTCCTGATCCCGAAGCTTTGCAACAGTTTACCATCTGGGTGGGT 124  
Db 140 GGAGCATCTCGCCACTTGATTCCTCGCTATGCGAGCAATCTATCATTTGGGATGGGT 199  
QY 125 CACTGGCACTGAGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTCCCTC 184  
Db 200 GACCGGCAACAGAGGTGGCATGAGCTTAAGTACACGATGAGATCTACATAGCACCCGTC 259  
QY 185 AGCGGTGCAAGAGGAGCGCATTCAGCCAGACAGCATGTTTGTGTGACATTAAATGAGCA 244  
Db 260 GGGCGTCCAGAGGAGCGAATGAGCGCGGAGGATCTCTTCGTCAGGATATTAACCGCAA 319  
QY 245 GGACATAAGCGGCTCCAGCATCTAAGAGCTGAAGAAAGCCAGTGCACCTCTCTTTT 304  
Db 320 GGATCTGCACTGCCCTTGAGATCAAGGGCTGAAGAGAGCCATGTACCGCGCTCTT 379  
QY 305 CATGAATGTTATACCATGAGAGGAGTGGCGCAGTGATTCATCCACTCTAAAGCTGC 364  
Db 380 TATGCTGGCTATCAGCATCGGCGGCGGAGCGCTCATCCACACCCATTGCGAGCAGC 439  
QY 365 TGTGATGGCTACCTTCTGTTCCAGGACAGGAGTTTAAATTTACACATCAAGAGATGAT 424  
Db 440 CGTAATGGCCACGCTCTGTGGCCAGGAAACCTTCGCGTCACCCACTTGGAGATGAT 499  
QY 425 CAAAGGAATAGGAATGTACCTCAGGAGCTATTACAGATACGATGATGTTAGTGGT 484  
Db 500 CAAGGGCTCTACGATGAGCGGACAGCATATTTGGCTACGAGGAGCTGCTGCT 559  
QY 485 ACCTATTATTGAGAACTCTCTGAGAGAGAGATCTCAAGAAAGAGATGCTCATGCCAT 544  
Db 560 ACCGATCATCGAAGACACACCTTTTGAACGCGACCTGGCGCAGATGATGACCGCCCAT 619  
QY 545 GAATGAGTACCGAGCTCTGTTGGGTTCTTTCGCGCTCATGGGTGTACGTTGGGG 604  
Db 620 GATGGAGTATCCGGGCTGCGAGTCCTGTTGTCGACGACGCGGCTCTACGTTGGGG 679  
QY 605 AGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACATTGC 664  
Db 680 ACAGACTGGAGAGAGCCAAACCATGTGGATGTCTATGACTATCTCTTCTCCATTGC 739  
QY 665 TGTCTCCATGAAGAGATGGGACTCGATCCAAACAGCTC 704  
Db 740 CGTGGAAATGAGAGCGCGGATCGATCCGAAAGTTC 779

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 22:00:56 ; Search time 96 Seconds  
(without alignments)  
4855.820 Million cell updates/sec

Title: US-09-937-905-1  
Perfect score: 840  
Sequence: 1 gaaccatgtctgctgtca.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323	38.5	409	4	US-09-621-976-9280
2	143	17.0	229	1	US-08-840-683-4
3	143	17.0	229	2	US-08-555-722-4
4	143	17.0	229	3	US-09-384-301-4
5	53.6	6.4	1223	3	US-09-154-874-4
6	53.6	6.4	1223	4	US-08-931-668-4
7	53.6	6.4	1223	4	US-09-468-175-4
8	53	6.3	147	4	US-09-621-976-10254
9	53	6.3	7218	1	US-08-232-463-14
10	52.8	6.3	997	4	US-09-907-794A-376
11	52.8	6.3	997	4	US-09-905-125A-376
12	52.8	6.3	997	4	US-09-902-775A-376
13	52.2	6.2	147	4	US-09-621-976-10383
14	50.4	6.0	375	3	US-08-946-026-23
15	49.6	5.9	242	4	US-09-621-976-16320
16	48.8	5.8	1074	3	US-09-248-335-67
17	48.6	5.8	746	3	US-09-013-810-1
18	48.6	5.8	1558	1	US-08-455-550-7
19	48.2	5.7	350	1	US-08-171-385-14
20	48.2	5.7	350	3	US-08-361-441B-14
21	48.2	5.7	5555	1	US-08-484-438-3
22	47.8	5.7	1987	4	US-09-227-357-44
23	47.6	5.7	441	4	US-09-601-537-10
24	47.6	5.7	674	4	US-09-620-405B-465
25	47.6	5.7	674	4	US-09-433-826B-465
26	47.6	5.7	674	4	US-09-604-287A-465
27	47.6	5.7	674	4	US-09-834-759-465

28	47.6	5.7	1454	4	US-09-372-422A-19	Sequence 19, Appl
29	47.6	5.7	2186	4	US-09-360-545-66	Sequence 66, Appl
30	47.6	5.7	4121	4	US-09-601-537-9	Sequence 9, Appl
31	47.4	5.6	1474	3	US-08-821-994-64	Sequence 64, Appl
32	47.4	5.6	2422	4	US-09-369-247-53	Sequence 53, Appl
33	47.2	5.6	1323	3	US-09-413-452-3	Sequence 3, Appl
34	47.2	5.6	1323	3	US-09-413-068-3	Sequence 3, Appl
35	47.2	5.6	1323	4	US-09-403-345A-3	Sequence 3, Appl
36	47.2	5.6	1700	2	US-08-897-340-4	Sequence 4, Appl
37	47.2	5.6	1700	3	US-09-252-329-4	Sequence 4, Appl
38	47.2	5.6	2434	4	US-09-489-847-67	Sequence 67, Appl
39	47.2	5.6	3715	4	US-09-234-245-1	Sequence 1, Appl
40	47	5.6	1332	2	US-09-057-762-1	Sequence 1, Appl
41	47	5.6	1332	3	US-08-326-119A-1	Sequence 1, Appl
42	47	5.6	2628	1	US-08-143-219-1	Sequence 1, Appl
43	46.8	5.6	193	4	US-09-621-976-10543	Sequence 10543, A
44	46.8	5.6	1069	4	US-09-205-258-74	Sequence 74, Appl
45	46.8	5.6	3124	4	US-09-734-030-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-621-976-9280  
; Sequence 9280, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PK2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 9280  
; LENGTH: 409  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-9280

Query Match	38.5%	Score 323;	DB 4;	Length 409;
Best Local Similarity	88.6%	Pred. No. 9.8e-84;		
Matches	350;	Conservative	0;	Mismatches 45;
			Indels	0;
			Gaps	0;
Qy	342	ATTCAATCCCACTCTAAAGCTGCTGATGGCTACCTTCTGTTTCCAGGACAGGAGTTT	401	
Db	1	ATTCAATCCCACTCTAAAGCTGCTGATGGCTACCTTCTGTTTCCAGGACAGGAGTTT	60	
Qy	402	AAATTTACATCAAGAGATGATCAAAAGGAATAGGAATGTACTCTCAGGAGGCTATTAC	461	
Db	61	AAATTTACATCAAGAGATGATCAAAAGGAATAGGAATGTACTCTCAGGAGGCTATTAT	120	
Qy	462	AGATACGATGATAGTTAGTGTACTTATTGAGAACACTCTCTGAGAGAGAGATCTC	521	
Db	121	AGATATGATGATAGTTAGTGTACTTATTGAGAACACTCTCTGAGAGAGAGATCTC	180	
Qy	522	AAAGAAAGGATGGCTCATGCCAATGAGTACCCAGACTCTCTGCGGTTCTTGTTCGG	581	
Db	181	AAAGATGAGATGGCTCATGCCAATGAGTACCCAGACTCTCTGCGGTTCTTGTTCGG	240	
Qy	582	CGTATGCGGTGTACGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT	641	
Db	241	CGTATGCGGTGTACGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT	300	
Qy	642	TATGACTACCTGTTTGTGATGCTGCTTCCATGAAGAGATGGGACTCGATCCACACAG	701	
Db	301	TATGACTATTTTGTGATGCTGCTTCCATGAAGAGATGGGACTCGATCCACACAG	360	
Qy	702	CTCCAGCTGGAGAAATGGAATGTCTTAAGCCAA	736	

Db 361 CTCCAGTTGGAGAAATGGAATTGTCTAAGCCAA 395

## RESULT 2

US-08-840-683-4  
; Sequence 4, Application US/08840683  
; Patent No. 5821051  
; GENERAL INFORMATION:  
; APPLICANT: Androphy, Elliot J.  
; APPLICANT: Chen, Jason J.  
; TITLE OF INVENTION: E6-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII (text)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/840,683  
; FILING DATE: 29-APR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/273,059  
; FILING DATE: 08-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: NEP-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 229 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA

US-08-840-683-4

Query Match 17.0%; Score 143; DB 1; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 1; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

APPLICANT: Androphy, Elliot J.  
APPLICANT: Chen, Jason J.  
TITLE OF INVENTION: E6-BINDING PROTEINS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/555,722  
FILING DATE: 14-NOV-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: NEP-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;

Best Local Similarity 85.2%; Pred. No. 7.3e-32;

Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108

Db 4 GCGGCCGACAGGACAGGACATCCAGATACCTGATCCAGAAC-TTGCAACAGTTTT 62

QY 109 ACCATCTGGGCTGGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168

Db 63 ACCATTTAGGCTGGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120

QY 169 TCTACATTGCTCCCTCAGCGTGCATAAGAGGAGCGATTCAGCCAGACATGTTTGTGT 228

Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAAAGCAATTCAGCTTGAAGCATGTTTGT 179

QY 229 GTGACATTATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277

Db 180 GTGATATAATGAAGAGGACATAAGTGAGACCTTCGCCATCGAAGAGCT 228

1 3538  
COMPUTER READABLE FORM:  
MEDIUM TYPE: PC SKITTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/154,874  
FILING DATE:  
CLASSIFICATION:

SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..1223
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..990
; US-08-931-668-4
;
; Query Match
; Best Local Similarity 73.9%; Score 53.6; DB 4; Length 1223;
; Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
;
; QY 749 GCATCTCCAACTAAACAACTCAATTATGCTTAAATAAATAAAGCTGCTTTTAAA 808
; DB 1113 GGATCTTGAATGTGAACGAGTTTAACTTTATCATTTAATTAAGCTCTGTTGTTTAAA 1172
;
; QY 809 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
; DB 1173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1204
;
; RESULT 7
; US-09-468-175-4
; Sequence 4, Application US/09468175
; Patent No. 6617493
; GENERAL INFORMATION:
; APPLICANT: FADER, GARY MICHAEL
; TITLE OF INVENTION: ISOFALVONE BIOSYNTHETIC ENZYMES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/468,175
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/931,668
; FILING DATE: SEPTEMBER 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1098-A
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..1223
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..990
;
; US-09-468-175-4
; Query Match
; Best Local Similarity 6.4%; Score 53.6; DB 4; Length 1223;
; Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
;
; QY 749 GCATCTCCAACTAAACAACTCAATTATGCTTAAATAAATAAAGCTGCTTTTAAA 808
; DB 1113 GGATCTTGAATGTGAACGAGTTTAACTTTATCATTTAATTAAGCTCTGTTGTTTAAA 1172
;
; QY 809 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
; DB 1173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1204
;
; RESULT 8
; US-09-621-976-10254
; Sequence 10254, Application US/09621976
; Patent No. 8639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10254
; LENGTH: 147
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-10254
;
; Query Match
; Best Local Similarity 6.3%; Score 53; DB 4; Length 147;
; Matches 86; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
;
; QY 730 AGCTCCCGAGTTGGAGAAAAATGGAATTGTTAAGCCAAAGTGGATGCTTAAGCATCTCCAAC 759
; DB 1 AACTCTGTTTGTAGAGAGATCTCAATGTTTGTAGTAAAGTACAAACACCATCAAAAAA 60
;
; QY 760 AATAAACAACTCAATTATGCTTAAATAAATAAAGCTGCTTTTAAAAA 819
; DB 61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120
;
; QY 820 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
; DB 121 AAAAAAAAAAAAAAAAAAAAAA 141
;
; RESULT 9
; US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-3300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZ9pt-Fls

US-08-232-463-14

Query Match 6.3%; Score 53; DB 1; Length 7218;

Best Local Similarity 7.0%; Pred. No. 4.6e-05;

Matches 29; Conservative 212; Mismatches 172; Indels 0; Gaps 0;

QY 428 AGGATAAGGAATGTACCTCAGGAGGCTTATACAGATACGATGATGTTAGTGTACC 487

DB 1493 AGGCATCACTGTGTAATTAATCTATGCAAGTAGTTAAAGAGATAGAAATTTGTGTA 1434

QY 488 TATTATTAGAACACTCTCAAGAGAGAGATCTCAAGAGAGAGATCGCTCATGCCATGAA 547

DB 1433 RRR 1374

QY 548 TGAGTACCAGACTCTGTGCGGTTCTTGTGCGGCTCATGCGGTGTACGTGGGAGA 607

DB 1373 RRR 1314

QY 608 AACATGGGAGAGAAACCAACATGTGAGTGTATGACTCTTTGACATTCCTGT 667

DB 1313 RRR 1254

QY 668 CTCATGAAGAGATGGGACTCGATCCACACAGCTCCGAGTTGGAGAAATGAT 727

DB 1253 RRR 1194

QY 728 GTAAGCAAGTGTAGCTTAAGCATCTCCAACTAAACAACTCAATATGCTTAA 787

DB 1193 RRR 1134

QY 788 TAAACTCAGCTCTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 840

DB 1133 RRR 1081

RESULT 10

US-09-907-794A-376

Sequence 376, Application US/09907794A

Patent No. 6635468

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/907,794A

CURRENT FILING DATE: 2001-07-17

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/23089

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: PCT/US99/28214

PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: PCT/US99/28313

PRIOR FILING DATE: 1999-11-30

PRIOR APPLICATION NUMBER: PCT/US99/28564

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/28565

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/30095

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: PCT/US99/30911

PRIOR FILING DATE: 1999-12-20

PRIOR APPLICATION NUMBER: PCT/US99/30999

PRIOR FILING DATE: 1999-12-20

PRIOR APPLICATION NUMBER: PCT/US00/00219

PRIOR FILING DATE: 2000-01-05

NUMBER OF SEQ ID NOS: 423

SEQ ID NO 376

LENGTH: 997

TYPE: DNA

ORGANISM: Homo Sapien

US-09-907-794A-376

Query Match 6.3%; Score 52.8; DB 4; Length 997;

Best Local Similarity 56.2%; Pred. No. 1.8e-05;

Matches 99; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 665 TGTCTCCATGAAGAGATGGGACTCCGATCCACACAGCTCCCGAGTTGGAGAAATGGAAT 724

DB 789 TGTAAAG 848

QY 725 TGTGTAAGCAAGTGGATGCTTAAGCATCTCCACATAAACAATTAATATGCTT 784

Db 849 GCTGAACCTCAAGGACTTCATAGTATTGTTACTCTGATACAAATAAAAAATAAGTAGT 908  
Qy 785 AAATAAACTCAGCTGCTTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 840  
Db 909 TTTAAATGTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 964

## RESULT 11

US-09-905-125A-376  
; Sequence 376, Application US/09905125A  
; Patent No. 6664376  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,125A  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08  
; PRIOR APPLICATION NUMBER: PCT/US99/20944  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: PCT/US99/21090  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/21547  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: PCT/US99/23089  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: PCT/US99/28214  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: PCT/US99/28313  
; PRIOR FILING DATE: 1999-11-30  
; PRIOR APPLICATION NUMBER: PCT/US99/28564  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/28565  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: PCT/US99/30095  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: PCT/US99/30911  
; PRIOR FILING DATE: 1999-12-20

; PRIOR APPLICATION NUMBER: PCT/US99/30999  
; PRIOR FILING DATE: 1999-12-20  
; PRIOR APPLICATION NUMBER: PCT/US00/00219  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 423  
; SEQ ID NO 376  
; LENGTH: 997  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-09-905-125A-376

Query Match 6.3%; Score 52.8; DB 4; Length 997;  
Best Local Similarity 56.2%; Pred. No. 1.8e-05;  
Matches 99; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
Qy 665 TGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCAAGTTGGAGAAAATGGAAT 724  
Db 789 TGTAAAGAAGAGAGCAAGACATGSCCAACCCCGCCACACAGAGAAATTTGTGC 848  
Qy 725 TGTGTAGCCCAAGTGGATGCTTAAGCATCTCCAAACATAAAACAACTCAATTATGCTT 784  
Db 849 GCTGAACCTTCAAGGACTTCATAGTATTGTTACTCTGATACAAATAAAAAATAAGTAGT 908  
Qy 785 AAATAAACTCAGCTGCTTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 840  
Db 909 TTTAAATGTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 964

## RESULT 12

US-09-902-775A-376  
; Sequence 376, Application US/09902775A  
; Patent No. 6686451  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/902,775A  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222  
; PRIOR FILING DATE: 1999-07-28  
; PRIOR APPLICATION NUMBER: PCT/US99/20594  
; PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30959  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 376  
LENGTH: 997  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-902-775A-376

Query Match 6.3%; Score 52.8; DB 4; Length 997;  
Best Local Similarity 56.2%; Pred No. 1.8e-05;  
Matches 99; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 665 TGCTCCATGAAGATGGGACTCGATCCACACAGCTCCCGTGGAGAAATGGAAT 724  
DB 789 TGTAAGAAGGAGAGCAAGCATGTGCCACCCACCCGCGCCACACGAGAAATTTGTGC 848  
QY 725 TGTGTAGCCCAAGTGGATGCCCTTAAGCATCTCCACAAATAAACAACCTCAATTATGCTT 784  
DB 849 GCTGAATCTCAAGGACTTCATAGTATTGTCTCTGATACAAATAAATAAGTAGT 908  
QY 785 AAATAAACTCAGCTGCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 840  
DB 909 TTTAATGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 964

RESULT 13  
US-09-621-976-10383  
Sequence 10383, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 10383  
LENGTH: 147  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-621-976-10383

Query Match 6.2%; Score 52.2; DB 4; Length 147;  
Best Local Similarity 59.8%; Pred. No. 9.8e-06;  
Matches 84; Conservative 2; Mismatches 55; Indels 0; Gaps 0;  
QY 700 AGCTCCAGTTGGAGAAATGGAATTGTGTGAAGCCAAAGTGGATGCCCTAAGCATCTCCAAC 759

DB 1 AACTCTGTTTGTAGAGAAGATCTCAATGTTTGTAGTAAAGTACAAACACCATCAAAAAA 60  
QY 760 AATAAAACAAACTCAATTATGCGCTTAATAAATAAATAAATAAATAAATAAATAA 819  
DB 61 RRAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 120  
QY 820 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
DB 121 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 141

RESULT 14  
US-08-946-026-23  
Sequence 23, Application US/08946026  
Patent No. 6034218  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Mitcham, Jennifer L.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946.026  
FILING DATE: 07-OCT-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/POCKET NUMBER: 210121.424C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-946-026-23

Query Match 6.0%; Score 50.4; DB 3; Length 375;  
Best Local Similarity 83.8%; Pred. No. 5.3e-05;  
Matches 57; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 773 CAATTATGCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 832  
DB 281 CCATTACAGCCCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 340  
QY 833 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840  
DB 341 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 348

RESULT 15  
US-09-621-976-16320  
Sequence 16320, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 22:19:21 ; Search time 597 Seconds  
(without alignments)  
6509.501 Million cell updates/sec

Title: US-09-937-905-1

Perfect score: 840

Sequence: 1 gaaccatgtctggtgtca.....aaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Published Applications NA.\*
- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
  - 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
  - 16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
  - 17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
  - 18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
  - 19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	840	13	US-10-381-710-1
2	721.2	85.9	726	17	US-10-629-329A-3
3	598.6	71.3	1183	16	US-10-220-381-28
4	595.4	70.9	1136	13	US-10-381-710-3
5	572.2	68.1	729	17	US-10-629-329A-1
6	399.8	47.6	3784	14	US-10-044-090-252
7	378.6	45.1	402	9	US-09-933-797-133
8	279.4	33.3	384	15	US-10-796-692-4183
9	279.4	33.3	384	15	US-10-040-862-4183
10	279.4	33.3	384	16	US-10-057-475B-4183
11	279.4	33.3	384	16	US-10-154-884B-4183
12	246.4	29.3	479	10	US-09-918-995-16268
13	192.4	22.9	468	13	US-10-085-783A-34366
14	192.4	22.9	468	16	US-10-242-535A-34366

15	130.4	22.7	664	15	US-10-106-698-871	Sequence 871, App
16	150.2	17.9	473	13	US-10-085-783A-50536	Sequence 50536, A
17	150.2	17.9	473	16	US-10-242-535A-50536	Sequence 50536, A
18	145	17.3	532	15	US-10-029-386-2402	Sequence 2402, Ap
19	141.2	16.8	194	15	US-10-029-386-16102	Sequence 16102, A
20	117.2	14.0	1541	13	US-10-425-114-13586	Sequence 13586, A
21	117.2	14.0	1656	13	US-10-425-114-13377	Sequence 33377, A
22	117.2	14.0	1897	13	US-10-425-114-15898	Sequence 15898, A
23	115	13.7	1765	13	US-10-425-114-133005	Sequence 33005, A
24	113.6	13.5	792	13	US-10-424-599-14328	Sequence 14328, A2
25	110.4	13.1	1830	13	US-10-425-114-12739	Sequence 12739, A
26	110.4	13.1	1886	13	US-10-424-599-130875	Sequence 130875, A
27	82.6	9.8	312	10	US-09-991-936-522	Sequence 522, App
28	82	9.8	433	9	US-09-954-456-1894	Sequence 1894, Ap
29	65	7.7	65	10	US-09-908-375-23950	Sequence 23950, A
30	64.2	7.6	255	16	US-10-131-827-8206	Sequence 8206, Ap
31	56.8	6.8	5884	15	US-10-311-455-529	Sequence 529, App
32	56.8	6.8	5884	17	US-10-257-166-39	Sequence 39, Appl
33	52.8	6.3	997	9	US-09-909-320-376	Sequence 376, App
34	52.8	6.3	997	9	US-09-909-088B-376	Sequence 376, App
35	52.8	6.3	997	9	US-09-905-291A-376	Sequence 376, App
36	52.8	6.3	997	9	US-09-902-853-376	Sequence 376, App
37	52.8	6.3	997	9	US-09-907-824-376	Sequence 376, App
38	52.8	6.3	997	9	US-09-907-841-376	Sequence 376, App
39	52.8	6.3	997	10	US-09-904-011-376	Sequence 376, App
40	52.8	6.3	997	10	US-09-906-742-376	Sequence 376, App
41	52.8	6.3	997	10	US-09-906-838-376	Sequence 376, App
42	52.8	6.3	997	10	US-09-907-613-376	Sequence 376, App
43	52.8	6.3	997	10	US-09-907-942-376	Sequence 376, App
44	52.8	6.3	997	10	US-09-904-859-376	Sequence 376, App
45	52.8	6.3	997	10	US-09-909-204-376	Sequence 376, App

ALIGNMENTS

RESULT 1

US-10-381-710-1  
; Sequence 1, Application US/10381710  
; Publication No. US20040052789A1  
; GENERAL INFORMATION:  
; APPLICANT: SHA, Shiken et al.  
; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME  
; FILE REFERENCE: 0230-0198P  
; CURRENT APPLICATION NUMBER: US/10/381,710  
; CURRENT FILING DATE: 2003-09-16  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Mouse macrophage cell RAW 264.7  
US-10-381-710-1

Query Match 100.0%; Score 840; DB 13; Length 840;  
Best Local Similarity 100.0%; Pred. No. 3.4e-210;  
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAACCATGTCGCTGCTCAAGCTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG	60
Db	1	GAACCATGTCGCTGCTCAAGCTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG	60
QY	61	ACAAGGAGCACCCCGGATTCCTGATCCCAAGAACTTTGCAACAGTTTACCAATCTGGGCT	120
Db	61	ACAAGGAGCACCCCGGATTCCTGATCCCAAGAACTTTGCAACAGTTTACCAATCTGGGCT	120
QY	121	GGTCACTGGCACTGGAGGGGAAATCAGCTTGAAGCATGGCAATGAATCTACATTCCTC	180
Db	121	GGTCACTGGCACTGGAGGGGAAATCAGCTTGAAGCATGGCAATGAATCTACATTCCTC	180
QY	181	CCTCAGCGCTGCAAAAGGAGCCATTCACCCAGACACATGTTTGTGTGACATTAATG	240
Db	181	CCTCAGCGCTGCAAAAGGAGCCATTCACCCAGACACATGTTTGTGTGACATTAATG	240

```
QY 241 AGCAGGACATAAGCGGSCCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 300
Db 241 AGCAGGACATAAGCGGSCCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 300
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Db 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 420
Db 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 420
QY 421 TGATCAAAAGGAATTAAGAAATGTACCTCAGGAGGCTATTACAGATAGCATATGTTAG 480
Db 421 TGATCAAAAGGAATTAAGAAATGTACCTCAGGAGGCTATTACAGATAGCATATGTTAG 480
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 540
Db 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 540
QY 541 CCATGAATGATGACCCAGACTCCTGTGCGGTTCTGTCGCGGTCATGGGCTGACGTGT 600
Db 541 CCATGAATGATGACCCAGACTCCTGTGCGGTTCTGTCGCGGTCATGGGCTGACGTGT 600
QY 601 GGGGAGAAACATGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 660
Db 601 GGGGAGAAACATGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 660
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACAGCTCCAGTCCAGTGGAGAAATG 720
Db 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACAGCTCCAGTCCAGTGGAGAAATG 720
QY 721 GAATGTGTAGCCAGAGTGGATGCTAGCATCTCCACATAAACAACTCAATTATG 780
Db 721 GAATGTGTAGCCAGAGTGGATGCTAGCATCTCCACATAAACAACTCAATTATG 780
QY 781 CCTTAATAATAAATACTCAGTCTGCTTTTAAAAAATAAATAAATAAATAAATAAATAA 840
Db 781 CCTTAATAATAAATACTCAGTCTGCTTTTAAAAAATAAATAAATAAATAAATAAATAA 840

RESULT 2
US-10-629-329A-3
; Sequence 3, Application US/10629329A
; Publication No. US20040086848A1
; GENERAL INFORMATION:
; APPLICANT: DARNAY, BRYANT G.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND
; FILE REFERENCE: UTSC:761US
; CURRENT APPLICATION NUMBER: US/10/629,329A
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: 60/399,205
; PRIOR FILING DATE: 2002-07-29
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-629-329A-3
Query Match 85.9%; Score 721.2; DB 17; Length 726;
Best Local Similarity 99.6%; Pred. No. 5.1e-179;
Matches 723; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 6 ATGCTGGCTGTCAAGCTCAAGAGACTGTTGCTCGCGCGCTGTCGCGGAGGACAAG 65
Db 1 ATGCTGGCTGTCAAGCTCAAGAGACTGTTGCTCGCGCGCTGTCGCGGAGGACAAG 60
```

```
QY 66 GAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCTGGTTC 125
Db 61 GAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCTGGTTC 120
QY 126 ACTGGCATTGGAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATTGCTCCCTCA 185
Db 121 ACCGGCATTGGAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATTGCTCCCTCA 180
QY 186 GGGCTGCAAAAGGAGCGCATTCAGCCAGAGAGACATGTTTGTGTGTGACATTAATGAGCAG 245
Db 181 GGGCTGCAAAAGGAGCGCATTCAGCCAGAGAGACATGTTTGTGTGTGACATTAATGAGCAG 240
QY 246 GACATAAGCGGGCTCCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCTTTTC 305
Db 241 GACATAAGCGGGCTCCTCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCTTTTC 300
QY 306 ATGAATGCTTATACCATGAGAGAGCTGGCGAGTGATTCATACCCACTCTAAAGCTGCT 365
Db 301 ATGAATGCTTATACCATGAGAGAGCTGGCGAGTGATTCATACCCACTCTAAAGCTGCT 360
QY 366 GTGATGGCTACCCCTTCTGTTTCCAGAGCAGGAGTTTAAAAATTACATCAAGAGATGATC 425
Db 361 GTGATGGCTACCCCTTCTGTTTCCAGAGCAGGAGTTTAAAAATTACATCAAGAGATGATC 420
QY 426 AAAGGAATAAGGAAATGATCCTCAGAGAGGCTATTACAGATAGCATGATATGTTAGTGA 485
Db 421 AAAGGAATAAGGAAATGATCCTCAGAGAGGCTATTACAGATAGCATGATATGTTAGTGA 480
QY 486 CCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGGCTCATGCCATG 545
Db 481 CCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGGCTCATGCCATG 540
QY 546 AATGAGTACCAGACTCCTGTGCGGTTCTTGTCCGGCTCATGGGGTGTAGTGTGGGA 605
Db 541 AACGAGTACCAGACTCCTGTGCGGTTCTTGTCCGGCTCATGGGGTGTAGTGTGGGA 600
QY 606 GAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGCT 665
Db 601 GAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGCT 660
QY 666 GTCTCCATGAAGAGATGGGACTCGATCCAAACAGACTCCAGTTGGAGAAATGGAATT 725
Db 661 GTCTCCATGAAGAGATGGGACTCGATCCAAACAGACTCCAGTTGGAGAAATGGAATT 720
QY 726 GTGTAA 731
Db 721 GTGTAA 726

RESULT 3
US-10-220-381-28
; Sequence 28, Application US/10220381
; Publication No. US20030207430A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BANDMAN, Olga
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; TITLE OF INVENTION: HUMAN ENZYME MOLECULES
; FILE REFERENCE: PP-0763 PCT
; CURRENT APPLICATION NUMBER: US/10/220,381
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NC 28
; LENGTH: 1183
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inceyte ID No. US20030207430A1 2116390CB1
US-10-220-381-28

Query Match      71.3%; Score 598.6; DB 16; Length 1183;
Best Local Similarity 84.9%; Pred. No. 1.1e-146;
Matches 694; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGCTCAAGCTC---AAGGAGACTGTGCTCGCGCGCTGTGGCGGCAGG 60
DB 92 CCATGCTGGCTGCTGATGCTCGGAGGAGACTGTGTTCCCGGAGATGCGGCGGCAGG 151
QY 61 ACAAGGAGACACCCCGATTCCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 120
DB 152 ACNAGGAGCATCCAGATACCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 211
QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 180
DB 212 GGGTCACTGGCACTGGAGGAGAAATAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 271
QY 181 CCTCAGGCTGCAAAAGAGCGCATTCAGCAGAGACATGTTGTGTGTCGATTAATG 240
DB 272 CTTTCAAGAGTGCAAAAGAGCAACGAATTCAGCCTGAAGCATGTTTGTGTGATATAATG 331
QY 241 AGCAGGACATAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCACCTCCTC 300
DB 332 AAGAAGCATAGTGAACCTTCGCCATCGNAGAGCTTAAAAAGCCAGTGTACTCCTC 391
QY 301 TTTTCATGAATGCTTATACATGAGAGGAGCTGGCGAGTGATTCATACCATCTTAAG 360
DB 392 TTTTCATGAATGCTTATACATGAGAGGAGCAGGTGAGTGTATCATACCATCTTAAG 451
QY 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGCAGGAGTTTAAATTTACATCAAGAGA 420
DB 452 CTGCTGTGATGGCGACCTTCTTCTTCCAGGAGCGGAGTTTAAATTTACATCAAGAGA 511
QY 421 TGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
DB 512 TGATAAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATGATGATGTTAG 571
QY 481 TGGTACTCTATTATGAGAACACCTCCTGAAGAGAGGATCTCAAGAGAAAGGATGGCTCATG 540
DB 572 TGGTACCCATATTATGAGATACACCTGAGGAGAAAGACCTCAAGATAGAAATGGCTCATG 631
QY 541 CCATGAATGATGACCGAGACTCCTGTGCGGTTCTTGTCCGCGCTCATGGGCTGACGTGT 600
DB 632 CAATGAATGAATACCCAGACTCCTGTGTCAGTCTGGTTCAGACGCTCATGGAGTATATGTT 691
QY 601 GGGGAGAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
DB 692 GGGGAGAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTATTTATTTTGATA 751
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACACAGCTCCCAAGTTGGGAAATG 720
DB 752 TTGCCGTATCAATGAAGAGATAGGACTTGTATCTTCCAGAGTCCCAAGTTGGGAAATG 811
QY 721 GAATTTGTGAAGCAAGTGATGCTTAAGCATCTCCAAAGTATTAAC-AAAACCTCAATTAT 779
DB 812 GAATTTGTGAAGCAAAAGAGAGTCTAATTTATATACAGATTAAGCTAAACGCTAATTAT 871
QY 780 GCCTTAATTAATCACTCAGCTGCTTTTAAAAAAGAAAA 816
DB 872 TATTTAAATGAAGCTATTTTTTTTAAATGAATGAAA 908

RESULT 4
US-10-381-710-3
; Sequence 3, Application US/10381710
; Publication No. US20040052789A1
; GENERAL INFORMATION:
; APPLICANT: SHA, Shiken et al.

; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME
; FILE REFERENCE: 0230-0198P
; CURRENT APPLICATION NUMBER: US/10/381,710
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-381-710-3

Query Match      70.9%; Score 595.4; DB 13; Length 1136;
Best Local Similarity 84.7%; Pred. No. 7.1e-146;
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGCTCAAGCTC---CAAGGAGACTGTGTTGCTCGCGCGCTGTGGCGGCAGG 60
DB 44 CCATGCTGGCTGCTGATGCTTGGAGGGAGACTGTGTTCCCGGAGATGCGGCGGCAGG 103
QY 61 ACAAGGAGACACCCCGATTCCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 120
DB 104 ACAAGGAGCATCCAGATACCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATTTAGGCT 163
QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 180
DB 164 GGGTCACTGGCACTGGAGGAGAAATAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 223
QY 181 CCTCAGGCTGCAAAAGAGAGCGCATTCAGCAGAGACATGTTTGTGTGTCGATTAATG 240
DB 224 CTTTCAAGAGTGCAAAAGAGCAACGAATTCAGCCTGAAGCATGTTTGTGTGATATAATG 283
QY 241 AGCAGGACATAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGTACTCCTC 300
DB 284 AAGAAGCATAGTGGAGCTTCGCCATCGAAGAGCTTAAAAAGCCAGTGTACTCCTC 343
QY 301 TTTTCATGAATGCTTATACATGAGAGGAGCTGGCGAGTGATTCATACCATCTTAAG 360
DB 344 TTTTCATGAATGCTTATACATGAGAGGAGCAGGTGAGTGTATCATACCATCTTAAG 403
QY 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGGAGCAGGAGTTTAAATTTACATCAAGAGA 420
DB 404 CTGCTGTGATGGCGACCTTCTTCTTCCAGGAGCGGAGTTTAAATTTACATCAAGAGA 463
QY 421 TGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
DB 464 TGATAAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATGATGATGTTAG 523
QY 481 TGGTACTCTATTATGAGAACACCTCCTGAAGAGAGGATCTCAAGAGAAAGGATGGCTCATG 540
DB 524 TGGTACCCATATTATGAGAAATACACCTTGAAGAGAAAGACCTCAAGAGATAGAAATGGCTCATG 583
QY 541 CCATGAATGATGACCCAGACTCCTGTGCGGTTCTTGTCCGCGCTCATGGGCTGTAOCTGT 600
DB 584 CAATGAATGAATACCCAGACTCCTGTGTCAGTACTGTGTGAGCTCATGGAGTATATGTT 643
QY 601 GGGGAGAACATGGGAGAGAAAGCAAAACCATGTGTGAGTGTATTAGTACTCCTGTTTGACA 660
DB 644 GGGGAGAACATGGGAGAGAGCCAAACCATGTGTGAGTGTATTAGTACTATTTATTTGATA 703
QY 661 TTGCTGTCTCCATGAAGAGATGGGAGTCTGATCCCAACACAGCTCCCAAGTTGGGAAATG 720
DB 704 TTGCCGTATCAATGAAGAGATAGGACTTGTATCTTCCAGAGTCCCAAGTTGGGAAATG 763
QY 721 GAATTTGTGAAGCAAGTGGATGCTTAAGCATCTCCAAAGTATTAAC-AAAACCTCAATTAT 779
DB 764 GAATTTGTGAAGCAAAAGAGAGTCTAATTTATATACAGATTAAGCTAAACGCTAATTAT 823
QY 780 GCCTTAATTAATCACTCAGCTGCTTTTAAAAAAGAAAA 816
DB 824 TATTTAAATGAAGCTATTTTTTTTAAATGAATGAAA 860
```

695 CACTCGATCCACACAGCTCCCGTGGAGAAATGGAAATTGTGTAA 731

Db 582 GTCCATGGAGTATATGTGTGGGGGAAACCAATGGGAGAGGCCAAAACCATGTGTGAGT 523  
QY 640 GTTATGACTTACTGTTTACACATTGCTCTCCATGAAGAAGTGGACTCGATCCATCCAAAC 699  
Db 522 GTTATGACTATTATTGATGATGCTGATGCGTATCAATGAAGAAGTAGGACTTGTATCCTTCC 463  
QY 700 AGCTCCCGAGTTGGAGAAATGGAATTTGTGAAGCAAGTGAATGCTTAAGCATCTCCAAAC 759  
Db 462 AGCTCCCGAGTTGGAGAAATGGAATTTGTGAAGCAAGTGAATGCTTAAGCATCTCCAAAC 403  
QY 760 AATAAAC--AACTCAATTATGCTTTAAATAAACTCAGCTGCTTTTAAAAAATAAAAA 817  
Db 402 GATAAAGCTAAACGTTAATTTATTTAAATGAAAGCTATTTTTTAAATGGAATGAAA 343

RESULT 7  
US-09-933-797-133/c  
; Sequence 133, Application US/09933797  
; Patent No. US20020155119A1  
; GENERAL INFORMATION:  
; APPLICANT: Robert A. Sikes et al.  
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital  
; TITLE OF INVENTION: Sinus Expressed Sequences  
; FILE REFERENCE: 9901-007-999  
; CURRENT APPLICATION NUMBER: US/09/933,797  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: US/09/482,933  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: PCT/US99/10746  
; PRIOR FILING DATE: 1999-05/14  
; PRIOR APPLICATION NUMBER: 60/085,383  
; PRIOR FILING DATE: 1998-05-14  
; NUMBER OF SEQ ID NOS: 811  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 133  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Murine  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(402)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-933-797-133

Query Match 45.1%; Score 378.6; DB 9; Length 402;  
Best Local Similarity 97.2%; Pred. No. 3.7e-89;  
Matches 385; Conservative 9; Mismatches 1; Indels 1; Gaps 1;  
QY 267 TCTAAGAAGCTGAAAAAAAGCCAGTGCACCTCTCTTTTCATGAATGCTTATACCATGAGA 326  
Db 402 TCTAAGAAGCTGAAWAAAAAGCCAGTGCACCTCTCTTTTCATGAATGCTTATACCATGAGA 343  
QY 327 -GGAGCTGGCGAGTGAATTCATACCCACTCTAAAGTGTGTGTGATGGCTACCTCTGT 385  
Db 342 GGGAGCTGGCGAGTGAATTCATACCCACTCTAAAGTGTGTGTGATGGCTACCTCTGT 283  
QY 386 TCCAGACAGGAGTTTAAATTAACATCAAGATGATCAAGAGGAATTAAGAAATGTAC 445  
Db 282 TCAGACAGGAGTTTAAATTAACATCAAGATGATCAAGAGGAATTAAGAAATGTAC 223  
QY 446 CTCAGGAGGCTATTACAGATACATGATGATGTTAGTGGTACCTATTATTGAGAACCTCC 505  
Db 222 CTCAGGAGGCTATTACAGATACATGATGATGTTAGTGGWACCDATTATTGAGAACCTCC 163  
QY 506 TGAAGAGAGGATCTCAAGAAAGGATGGCTATGCGATGAATGATGATGCCAGACTCTCG 565  
Db 162 TGAAGAGAGGATCTMAAGAAAGGATGGCTATGCGATGAATGATGATGCCAGACTCTCG 103  
QY 566 TGGGTTCTTGTCCGGTTCATGGGTGTACGTGTGGGAGAACATGGGAGAACGAA 625  
Db 102 TGCRTGTTCTTGCRRNRTCATGGGTGTACGTGTGGGAGAACATGGGAGAACGAA 43

QY 626 AACCATGTGTGAGTGTATGACTACCTACCTGTTTGACAT 661  
Db 42 AACCATGTGTGAGTGTATGACTACCTACCTGTTTGACAT 7

RESULT 8  
US-09-796-692-4183  
; Sequence 4183, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4183  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-796-692-4183

Query Match 33.3%; Score 279.4; DB 9; Length 384;  
Best Local Similarity 84.8%; Pred. No. 4.1e-63;  
Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 402 AAAATTAACATCAAGATGATCAAGGAATAAGAAATGTACCTCAGGAGCTATTATC 461  
Db 1 AAAATTAACATCAAGATGATCAAGGAATAAGAAATGTACCTCAGGAGCTATTAT 60  
QY 462 AGATACGATGATGTAGTGTGTACCTATTATTGAGAACACTCTCGAAGAGAGATCTC 521  
Db 61 AGATATGATGATGTAGTGTGTACCTATTATTGAGATACACTTGGAGAGAACCTC 120  
QY 522 AAAGAAAGATGGCTCATGCGCATGATGATGATGATGATGATGATGATGATGATGATG 581  
Db 121 AAAGATAGATGGCTCATGCAATGAATGAATACCCAGACTCTCTGTGCTAGTGTGTCAG 180  
QY 582 CGTGATGGGCTGTACGTGTGGGAGAACATGGGAGAACCAAAACCATGTGTGATGT 641  
Db 181 CGTGATGGAGTATATGTGTGGGGGAAACATGGGAGAGGCCAAACCATGTGTGATGT 240  
QY 642 TATGACTACCTCTTTGACATTTCTCTCATGAAGAGATGGGACTTCGATCCAAACACAG 701  
Db 241 TATGACTATTTTATTTGATATTTGCCGTATCAATGAAGAAAGTAGGACTTTCATCCTTCACAG 300

Db	121	AAAGATAGAA	TGGGCTCAT	GGCAATGA	ATACCCAGACT	CTCTGTGCAGT	ACTGCTCAG	180
Qy	582	CGTCATGGGG	GTGTACGTGT	GGGGAGAA	ACATGGGAGAA	ACAAAAACCA	ATGTGTGAGTGT	641
Db	181	CGTCATGGAG	TATATGTCT	GGGGGGAA	ACATGGGAGAA	GGCCAAAAAC	CATGTGTGAGTGT	240
Qy	642	TATGACTACCT	GTTTGACAT	TGCTGTCT	CCCATGGAAGAG	ATGGGACTCCG	ATCCAAACACAG	701
Db	241	TATGACTAT	TTTATTTGAT	TATTTCCGCGT	ATCAATGAAGAA	AGTAGGACTTTG	ATCTTCACAG	300
Qy	702	CTCCCAAGT	TGGAGAAA	ATGGAAATGT	GTAAAGCCAA	GTGGATCGCT	TAAGCATCTCCAA	761
Db	301	CTCCCAAGT	TGGAGAAA	ATGGAAATGT	CTAAAGCCAA	AGTAAGTCT	TAATTATATACAG	360
Qy	762	TAAACAAA	770					
Db	361	TAAAGCTAA	369					

## RESULT 10

RESULT 10  
US-10-057-475B-4183  
; Sequence 4183, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordonez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for  
; Treatment of Hematological Malignancies  
; FILE REFERENCE: 014058-01440205  
; CURRENT APPLICATION NUMBER: US/10/057.475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See  
; NUMBER CF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4183  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-4183

Query Match 33.3%; Score 279.4; DB 16; Length 384;  
Best Local Similarity 84.8%; Pred. No. 4.1e-63;  
Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 402 AAAATTACATCAAGAGATGATCAAAGGAATATAGGAATGTACCTCAGAGGCTATTAC 461

702	CTCCAGTGGAGAAAATGGAATTGTGAAGCCAAAGTGGATGCTTAGAGATCTCCACAA	761
QY		
Db		
301	CTCCAGTGGAGAAAATGGAATTGTGAAGCCAAAAGAAAGTCTAATTATATACAGAG	360
762	TAAACAAA	770
QY		
Db		
361	TAAAGCTAA	369

## RESULT 9

RESULT 9  
US-10-040-862-4183  
Sequence 4183, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
FILE REFERENCE: 014058-013520US  
CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIORITY APPLICATION NUMBER: US 60/186,126  
PRIORITY FILING DATE: 2000-03-01  
PRIORITY APPLICATION NUMBER: US 60/190,479  
PRIORITY FILING DATE: 2000-03-17  
PRIORITY APPLICATION NUMBER: US 60/200,545  
PRIORITY FILING DATE: 2000-04-27  
PRIORITY APPLICATION NUMBER: US 60/200,303  
PRIORITY FILING DATE: 2000-04-28  
PRIORITY APPLICATION NUMBER: US 60/200,779  
PRIORITY FILING DATE: 2000-04-28  
PRIORITY APPLICATION NUMBER: US 60/200,999  
PRIORITY FILING DATE: 2000-05-01  
PRIORITY APPLICATION NUMBER: US 60/202,084  
PRIORITY FILING DATE: 2000-05-04  
PRIORITY APPLICATION NUMBER: US 60/206,201  
PRIORITY FILING DATE: 2000-05-22  
PRIORITY APPLICATION NUMBER: US 60/218,950  
PRIORITY FILING DATE: 2000-07-14  
PRIORITY APPLICATION NUMBER: US 60/222,903  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: US 60/223,416  
PRIORITY FILING DATE: 2000-08-04  
PRIORITY APPLICATION NUMBER: US 60/223,378  
PRIORITY FILING DATE: 2000-08-07  
PRIORITY APPLICATION NUMBER: US 09/796,692  
PRIORITY FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4183  
LENGTH: 384  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-4183

	Query Match	33.3%;	Score 279.4;	DB 15;	Length 384;
	Best Local Similarity	84.8%;	Pred. No. 4.1e-63;		
	Matches 313;	Conservative 0;	Mismatches 56;	Indels 0;	Gaps 0;
QY	402	AAAATTACATCAAGAGATGATCAAGGAATTAAGGAATGTACTCTCAGGAGGCTATTAC	461		
Db	1	AAAATTACATCAAGAGATGATTAAGGAATTAAGGAATGTACTCTCCGAGGGTATTAT	60		
QY	462	AGATACGATGATATGTTAGTGGTACTATTATTGAGAACACTCTTGAAGAGAGAGATCTC	521		
Db	61	AGATATGATGATATGTTAGTGGTATCCCATTTATTGAGAATACACCTGAGGAGAGAACCTC	120		
QY	522	AAAGAAAGGATGGCTCATGCCATGAATGAGTACCCAGACTCTCTGTGCGGTTCTTCTCGG	581		

```
Db 1 AAAATTACACATCAAGAGATGATAAAAGGAATAAAGAAATGTAATCTTCGGAGGGTATTAT 60
QY 462 AGATACGATGATATCTTGTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 521
Db 61 AGATATGATGATATGTTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 120
QY 522 AAAGAAAGGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 581
Db 121 AAAGTAAGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 582 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 641
Db 181 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 240
QY 642 TATGACTACCTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
Db 241 TATGACTATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 702 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 761
Db 301 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 360
QY 762 TAAACAAA 770
Db 361 TAAAGCTAA 369

RESULT 11
US-10-154-884B-4183
; Sequence 4183, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4183
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-4183

Query Match 33.3%; Score 279.4; DB 16; Length 384;
Best Local Similarity 84.8%; Pred. No. 4.1e-63;

Db 1 AAAATTACACATCAAGAGATGATAAAAGGAATAAAGAAATGTAATCTTCGGAGGGTATTAT 60
QY 462 AGATACGATGATATCTTGTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 521
Db 61 AGATATGATGATATGTTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 120
QY 522 AAAGAAAGGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 581
Db 121 AAAGTAAGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 582 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 641
Db 181 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 240
QY 642 TATGACTACCTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
Db 241 TATGACTATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 702 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 761
Db 301 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 360
QY 762 TAAACAAA 770
Db 361 TAAAGCTAA 369

RESULT 12
US-09-918-995-16268
; Sequence 16268, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16268
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(479)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16268

Query Match 29.3%; Score 246.4; DB 10; Length 479;
Best Local Similarity 85.4%; Pred. No. 2.1e-54;
Matches 287; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

QY 4 CCATGCTCGGCTGTCAAGCT---CAAGGAGACTGTTGCTCGGGCCGTTGCGGCGCAGG 60
Db 144 CCATGCTCGGCTGTGATGCTTTGGGAGGAGACTGTTGTTCCCGAGATGCGGCGCGCAGG 203
QY 61 ACAAGGAGCACCCCGATTTCCTGATCCAGAACTTTGCAAAACAGTTTTTACCATCTGGGCT 120
Db 204 ACACGAGCATCCAGATACCTGATCCAGAACTTTGCAAAACAGTTTTTACCATTAGGCT 263
QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATGGAATCTACATGCTC 180
Db 264 GGGTCACTGGGACTGGAGAGGAATTAGCTTGAAGCATGGCGAATGGAATCTACATGCTC 323
QY 181 CCTCAGGGGTGCAAAAGGAGGCCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240

Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 402 AAAATTACACATCAAGAGATGATAAAAGGAATAAAGAAATGTAATCTTCAGGAGGGTATTAT 461
Db 1 AAAATTACACATCAAGAGATGATAAAAGGAATAAAGAAATGTAATCTTCAGGAGGGTATTAT 60
QY 462 AGATACGATGATATCTTGTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 521
Db 61 AGATATGATGATATGTTAGTGTACCTATTATTGAGAACACTCTCTGAGAGAAAGGATCTC 120
QY 522 AAAGAAAGGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 581
Db 121 AAAGTAAGATGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 582 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 641
Db 181 CGTATGGGTGTAGTGTGGGAGAACATGGGAGAACATGGGAGAACATGGGAGAACATGGGAG 240
QY 642 TATGACTACCTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
Db 241 TATGACTATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 702 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 761
Db 301 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCAAAGTGGGACTCGATCCCAACACAG 360
QY 762 TAAACAAA 770
Db 361 TAAAGCTAA 369
```



Db 324 CTCAGAGTGCCTTAAAGAGGACGCTTACGCTGAGACATGTTGTTATGATTAATG 383  
Qy 241 AGCAGGACATAAGCGGCTCCAGCATCTTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300  
Db 384 AAAAGGACATAAGTGGACCTTCGCTATCGAAGAGCTTAAAAAGGAGCCAGCTACTCCTC 443  
Qy 301 TTTTCATGATGCTTATACCATGAGGAGCTGGCG 336  
Db 444 TTTTCATGATGCTTATACCATGAGGAGCAGGTG 479

RESULT 13  
US-10-085-783A-34366  
; Sequence 34366, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34366  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6)..(6)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-085-783A-34366

Query Match 22.9%; Score 192.4; DB 13; Length 469;  
Best Local Similarity 78.6%; Pred. No. 3.2e-40;  
Matches 254; Conservative 0; Mismatches 67; Indels 2; Gaps 2;  
Qy 496 AGAACACTCTCTGAAGAGAGGATCTCAAG-AAAGATGGCTCATGCCATGATGATGATG 554  
Db 3 AGANACACCTCGAGGAGAGACCTCAAGATATAGATGGCTCATGCAATGAATGAATAC 62  
Qy 555 CCAGACTCTCTGCGGCTTCTTGTCCGGCTCATGGGCTGACGTGGGGGAGAAACATGG 614  
Db 63 CCAGACTCTCTGCGAGTACTGGTCAGACGTCTATGATATATGTTGGGGGAGAAACATGG 122  
Qy 615 GAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTGTGACATGCTGTCTCCATG 674  
Db 123 GAGAAAGCCAAACCATGTGTGAGTGTATGACTATTTATTTGATTTATTCGGTATCAATG 182  
Qy 675 AAGAGATGGAGCTCGATCCAAACACAGCTCCAGTGGAGAAATGGAATGTGTAAAGCC 734  
Db 183 AAGAAAGTAGGACTTGAATCTTCAAGCTCCAGTCCAGTGGAGAAATGGAATGTGTAAAGCC 242  
Qy 735 AAGTGGATGCTTACCATCTCCAAACATAAAC-AAACTCAATTATGCTTAAATAAALAC 793  
Db 243 AAAAGAAAGTCTAATATATATACAGAGATAAGCTAAACGTATATTTATTTAAATGAAG 302  
Qy 794 TCAGCTGCTTTTAAAAAATAAAAA 816  
Db 303 CTATTTTAAATGAATGAAA 325

RESULT 14  
US-10-242-535A-34366  
; Sequence 34366, Application US/10242535A

Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34366  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6)..(6)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-242-535A-34366

Query Match 22.9%; Score 192.4; DB 16; Length 468;  
Best Local Similarity 78.6%; Pred. No. 3.2e-40;  
Matches 254; Conservative 0; Mismatches 67; Indels 2; Gaps 2;  
Qy 496 AGAACACTCTCTGAAGAGAGGATCTCAAG-AAAGATGGCTCATGCCATGATGATGATG 554  
Db 3 AGANACACCTCGAGGAGAGACCTCAAGATATAGATGGCTCATGCAATGAATGAATAC 62  
Qy 555 CCAGACTCTCTGCGGCTTCTTGTCCGGCTCATGGGCTGACGTGGGGGAGAAACATGG 614  
Db 63 CCAGACTCTCTGCGAGTACTGGTCAGACGTCTATGATATATGTTATTTGATTTATTCGGTATCAATG 122  
Qy 615 GAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTGTGACATGCTGTCTCCATG 674  
Db 123 GAGAAAGCCAAACCATGTGTGAGTGTATGACTATTTATTTGATTTATTCGGTATCAATG 182  
Qy 675 AAGAGATGGAGCTCGATCCAAACACAGCTCCAGTGGAGAAATGGAATGTGTAAAGCC 734  
Db 183 AAGAAAGTAGGACTTGAATCTTCAAGCTCCAGTCCAGTGGAGAAATGGAATGTGTAAAGCC 242  
Qy 735 AAGTGGATGCTTACCATCTCCAAACATAAAC-AAACTCAATTATGCTTAAATAAALAC 793  
Db 243 AAAAGAAAGTCTAATATATATACAGAGATAAGCTAAACGTATATTTATTTAAATGAAG 302  
Qy 794 TCAGCTGCTTTTAAAAAATAAAAA 816  
Db 303 CTATTTTAAATGAATGAAA 325

RESULT 15  
US-10-106-698-871  
; Sequence 871, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide;  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03

Wed Jun 2 09:13:36 2004

```

; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 871
; LENGTH: 664
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)..(39)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-871

Query Match      22.7%; Score 190.4; DB 15; Length 664;
Best Local Similarity 79.0%; Pred. NO. 1.3e-39;
Matches 263; Conservative 0; Mismatches 67; Indels 3; Gaps 3;

QY 485 ACCATTATTGAGAACACCTCCTGAAGAGAGAGATCTCAAGAAAGGATGGCTCATGCCAT 544
Db 1 ACCATTATTGAGAAATACACCTGAGGAGAA-GACCTCAANGATAGWATGGCTCATGCAAT 59

QY 545 GAATGAGTACCCAGACTCCTGTGCGGTTCTTGTCCGGCGTCATGGGGTGACGTGTTGGG 604
Db 60 GAATGAATACCCAGACTCCTGTGCGGTTCTTGTCCGGCGTCATGGGGTGACGTGTTGGG 119

QY 605 AGAAACATGGGAGAGAAACCAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGC 664
Db 120 -GAAACATGGGAGAGAGAAACCAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGC 178

QY 665 TGCTCCATCAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATGGAAT 724
Db 179 CGTATCAATGAGAGAAAGTAGGACTTGATCCTTCACAGCTCCCGAGTTGGAGAAATGGAAT 238

QY 725 TGTGTAAAGCCCAAGTGGATGCGTAAAGCATCTCCAAACAATAAAC-AAACTCAATTATGCTT 783
Db 239 TGTCTAAGCCCAAGTGGATGCGTAAAGCATCTCCAAACAATAAAC-AAACTCAATTATGCTT 298

QY 784 TAAATAAACTCAGCTGCTTTTAAAAA 816
Db 299 TAAATGAAGCTATTTTAAAAATGAA 331
```

Search completed: June 2, 2004, 01:15:02  
Job time : 589 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 1, 2004, 21:54:22 ; Search time 3630 seconds  
(without alignments)  
6910.257 Million cell updates/sec

Title: US-09-937-905-1  
Perfect score: 840  
Sequence: 1 gaacacgtctggctgtcaa.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmd:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	92.9	897	11	AK077705 Mus muscu
2	741.2	88.2	939	13	BUS16668 AGENCOURT
3	687.2	81.8	704	13	BY756144 BY756144
4	674.2	80.3	739	14	CB595118 AGENCOURT

5	661.2	78.7	677	13	BY755126
6	641.8	76.4	645	9	AV306484
7	620	73.8	679	12	BG277254
8	618.6	73.6	642	10	AW98838
9	605.6	72.1	850	13	BUS58424
10	597	71.1	1201	9	AL525537
11	596.4	71.0	1034	13	BK362087
12	585.8	69.7	645	13	PQ193148
13	585	69.6	1124	9	AL559009
14	577	68.7	597	14	CF621082
15	574.6	68.4	710	13	BY735766
16	566.8	67.5	865	13	BUS21864
17	564.8	67.2	591	14	CD421847
18	563.8	67.1	591	10	BF608531
19	562	66.9	588	14	CF929450
20	561.6	66.9	1201	9	AL525491
21	558	66.4	770	14	CD521711
22	557.6	66.4	564	12	BG228307
23	557.2	66.3	580	10	BF019317
24	549.2	65.4	672	29	AY405038
25	546	65.0	602	9	AI172106
26	526.8	62.7	875	13	BUI82280
27	522.8	62.2	543	13	BX525605
28	517.8	61.6	600	12	BI985626
29	516	61.4	1102	9	AL580710
30	514.2	61.2	669	12	BM788265
31	510.6	60.8	714	12	BG722097
32	508.4	60.5	531	9	AL022608
33	507	60.4	572	13	BU760396
34	504.6	60.1	624	10	BB619396
35	503.6	60.0	738	12	BI185766
36	502.2	59.8	507	9	AI326904
37	490.2	58.4	552	9	AI171255
38	486.8	58.0	513	14	CD421948
39	486.4	57.9	810	12	BG209009
40	485.8	57.8	492	9	AA250391
41	485.8	57.8	923	10	BB871032
42	484.4	57.7	486	10	BB626187
43	480.8	57.2	657	12	BM762397
44	480.8	57.2	1461	10	BF970850
45	480.6	57.2	1004	12	BM450291

## ALIGNMENTS

## RESULT 1

## AK077705

## LOCUS

## DEFINITION

## AK077705

## AK077705.1

## GI:26346638

## HITC; CAP trapper.

## SOURCE

## Mus musculus

## ORGANISM

## REFERENCE

## Carninci, P. and Hayashizaki, Y.

## High-efficiency full-length cDNA cloning

## Meth. Enzymol. 303, 19-44 (1999)

## MEDLINE

## 99279253

## PUBMED

## 10349636

## REFERENCE

## Carninci, P. and Hayashizaki, Y.

## Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

## Normalization and subtraction of cap-trapper-selected cDNAs to

## prepare full-length cDNA libraries for rapid discovery of new genes

## Genome Res. 10 (10), 1617-1630 (2000)

## JOURNAL

## 20499374

## MEDLINE

## 11042159

AK077705 897 bp mRNA linear HTC 18-SEP-2003  
Mus musculus 8 days embryo whole body cDNA, RIKEN full-length  
enriched library, clone:5730540M11 product:monocyte macrophage 19,  
full insert sequence.

AK077705

AK077705.1

GI:26346638

HITC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374

MEDLINE

11042159

REFERENCE  
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed capillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

## TITLE

JOURNAL  
MEDLINE  
PUBMEDREFERENCE  
AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

## TITLE

JOURNAL  
REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 897)

JOURNAL  
REFERENCE  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanishi, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission

TITLE  
JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/

## FEATURES

## source

1. .897  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:5730540M11"  
/db\_xref="MGI:2393885"  
/db\_xref="taxon:10090"  
/clone="5730540M11"  
/tissue\_type="whole body"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="8 days embryo"  
115. .840  
/note="unlabeled protein product; monocyte macrophage 19 (MGD) MGI:1926788, GB|NM\_019735, evidence: BLASTN, 99%, match=786)  
putative"  
/codon\_start=1  
/protein\_id="BAC36968.1"  
/db\_xref="GI:26346639"

## CDS

/translation="MSGQAGDCSSRRPCGAQDKHPRFLIPELCKQFYHLNWTGAGSLKHGNEIYIAPSGVQKRIQPEDFVQDINEQDISGPPASKKLKKSQCTPLEFNAYTMRGAGAVIHTHSAAVNATILFPQEFKITHQEMIKRKTSGGYRYDDMLVPIITENPEKDLKRMAMHANNVPSCAVLVRRHVVVVGETWEKAKTMCEDYLFDIAYSMKMGGLDPTQLPVGENV"  
869. .874  
/note="putative"  
897  
/note="putative"

polyA\_signal

polyA\_site

## ORIGIN

Query Match 92.9%; Score 780; DB 11; Length 897;  
Best Local Similarity 99.4%; Pred. No. 2.5e-115;  
Matches 783; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTTGTCTCGCGCCGTGTGGCGCGCAGG 60  
DB 110 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTTGTCTCGCGCCGTGTGGCGCGCAGG 169  
QY 61 ACAAGGAGCACCCCGGATCTCTGATCCAGAACTTTGCAACAGCTTTTACCATCTGGGCT 120  
DB 170 ACAAGGAGCACCCCGGATCTCTGATCCAGAACTTTGCAACAGCTTTTACCATCTGGGCT 229  
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTC 180  
DB 230 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTC 289  
QY 181 CCTCAGGGCGTCCAAAGAGGCGCATTCAGCCAGAGAGACATCTTTGTGTGTGACATTAATG 240  
DB 290 CCTCAGGGCGTCCAAAGAGGCGCATTCAGCCAGAGAGACATCTTTGTGTGTGACATTAATG 349  
QY 241 AGCAGGACATAAGCGGGCGCTCAGCATCTAAGAGCTGAAAGAGGCGAGTGCATCTCTC 300  
DB 350 AGCAGGACATAAGCGGGCGCTCAGCATCTAAGAGAGTGAAGAGGCGAGTGCATCTCTC 409  
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGGAGTGTTCATACCCACTCTAAG 360  
DB 410 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGGAGTGTTCATACCCACTCTAAG 469  
QY 361 CTGCTGTGATGGTACCTCTCTGTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420  
DB 470 CTGCTGTGATGGTACCTCTCTGTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 529  
QY 421 TGATCAAGGAATAAGGAATAAGTACCTCAGGAGGCTATTACAGATCAGTATGTTAG 480  
DB 530 TGATCAAGGAATAAGGAATAAGTACCTCAGGAGGCTATTACAGATCAGTATGTTAG 589  
QY 481 TGGTACCTATTATTCAGACACTCTCTGAGAGAGAGTCTCAAGAGAGGATGGCTCATG 540  
DB 590 TGGTACCTATTATTCAGACACTCTCTGAGAGAGAGTCTCAAGAGAGGATGGCTCATG 649  
QY 541 CCATGAATGATACCCAGACTCTCTGCGGTTCTTTGTCGGCGTCAATGGGGTGTAGTGT 600  
DB 650 CCATGAATGATACCCAGACTCTCTGCGGTTCTTTGTCGGCGTCAATGGGGTGTAGTGT 709  
QY 601 GGGGAGAAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGA 660  
DB 710 GGGGAGAAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGA 769  
QY 661 TTGCTGTCTCCATGAAGAAGATGGGAGTGGAGTCCAGACAGCTCCCGAGTTGGAGAAATG 720  
DB 770 TTGCTGTCTCCATGAAGAAGATGGGAGTGGAGTCCAGACAGCTCCCGAGTTGGAGAAATG 829  
QY 721 GAATGTGTGAAGCAAGTGGATGCCTTAAGCATCTCCCAACATTAACCAAACTCAATTATG 780  
DB 830 GAATGTGTGAAGCAAGTGGATGCCTTAAGCATCTCCCAACATTAACCAAACTCAATTATG 889  
QY 781 CCTTAAT 788  
DB 890 CCTTAAGT 897

RESULT 2

BU516668	939 bp	mRNA	linear	EST 12-SEP-2002
LOCUS	AGENCOURT 10118491 NIH_MGC_134	Mus musculus	CDNA clone	
DEFINITION	IMAGE:6513424 5', mRNA sequence.			
ACCESSION	BU516668			
VERSION	BU516668.1	GI:22824206		
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 939)			
TITLE	NIH-MGC http://mgi.nci.nih.gov/.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: Dr. David Rowe			
	cDNA Library Preparation: Invitrogen Corp			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: LLAM14087 row: b column: 17			
	High quality sequence stop: 604.			
FEATURES	Location/Qualifiers			
source	1..939			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:6513424"			
	/tissue_type="undifferentiated limb"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_lib="NIH MGC 134"			
	/notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."			
ORIGIN				
Query Match	88.2%	Score 741.2;	DB 13;	Length 939;
Best Local Similarity	97.3%	Pred. No. 3.9e-109;		
Matches	786;	Conservative 0;	Mismatches 18;	Indels 4; Gaps 3;
QY	1	GAACCATGCTGCTCTCAAGCTCAGGAGACTGTTGCTCGCGCGCTGTGGCGCGCAGG	60	
DB	79	GAACCATGCTGCTCTCAAGCTCAGGAGACTGTTGCTCGCGCGCTGTGGCGCGCAGG	138	
QY	61	ACAAGGAGCACCCCGCTCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCT	120	
DB	139	ACAAGGAGCACCCCGCTCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCT	198	
QY	121	GGGTCTACGCTAGGAGGGGAATCAGTTGAAGATGCAATGAAATCAATGCTC	180	
DB	199	GGGTCTACGCTAGGAGGGGAATCAGTTGAAGATGCAATGAAATCAATGCTC	258	
QY	181	CCTGAGCGTGAAGAGGCGCATTCAGCAGACATGTTGTGTGATTAATG	240	
DB	259	CCTGAGCGTGAAGAGGCGCATTCAGCAGACATGTTGTGTGATTAATG	318	
QY	241	AGCAGGACATPAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCCAGTGCACTCCTC	300	
DB	319	AGCAGGACATPAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCCAGTGCACTCCTC	378	
QY	301	TTTTTCATGAATGCTTATPACCATGAGAGGAGCTGCGCAGTGATTCATPACCATCTAAAG	360	
DB	379	TTTTTCATGAATGCTTATPACCATGAGAGGAGCTGCGCAGTGATTCATPACCATCTAAAG	438	
QY	361	CTGCTGTGATGGTACCCCTCTGTTCCAGACAGGAGTTTAAATTTACATCATCAAGAGA	420	
DB	439	CTGCTGTGATGGTACCCCTCTGTTCCAGACAGGAGTTTAAATTTACATCATCAAGAGA	498	
QY	421	TGATCAAAAGGAATAGGAAATGTACCTCAGGAGGCTATTACAGATACGATATGTTAG	480	
DB	499	TGATCAAAAGGAATAGGAAATGTACCTCAGGAGGCTATTACAGATACGATATGTTAG	558	
QY	481	TGTTACCTATTATTGAGAACTCTCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG	540	
DB	559	TGTTACCTATTATTGAGAACTCTCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG	618	
QY	541	CCATCAATGAGTACCCAGACTCTCTGCGGTTCTTGTCCGGGCTCATGGGCTGAGTGT	600	
DB	619	CCATCAATGAGTACCCAGACTCTCTGCGGTTCTTGTCCGGGCTCATGGGCTGAGTGT	678	
QY	601	GGGAGAAACATGGGAGAAAGCAAAACCATCTGTGAGTGTATCACTACCTGTTGACA	660	
DB	673	GGGAGAAACATGGGAGAAAGCAAAACCATCTGTGAGTGTATCACTACCTGTTGACA	737	
QY	661	TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACACAGCTCCAGTTGAGAAATG	720	
DB	733	TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACACAGCTCCAGTTGAGAAATG	797	
QY	721	GAATTTGT-GTAAAGCAAGTGAATGCTTAAGCATCTCCCAACATAAAACCACTCAATT	777	
DB	799	GAATTTGTGAAGCCAGTGGATGCTTAAGCATCTCCCAACATAAAACCACTCAATT	857	
QY	778	ATGCTTTAAATAAACTCAGCTGCTTTT	805	
DB	858	TGCCCTTAATTAATAAACTCAGCTGCTTTT	885	
RESULT 3				
LOCUS	BY756144	704 bp	mRNA	linear
DEFINITION	BY756144 RIKEN full-length enriched, 1 cell embryo Mus musculus			
ACCESSION	BY756144	GI:27188993		
VERSION	EST.			
KEYWORDS	Mus musculus (house mouse)			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 704)			
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., Mckenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, J., Reid, J., Ring, B.Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Ullrich, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Harozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.			
TITLE	Analysis of the mouse transcriptome based on functional annotation			
JOURNAL	of 60,770 full-length cDNAs			
MEDLINE	Nature 420, 563-573 (2002)			
	22354683			

PUBMED  
COMMENT

12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp;  
URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Wataniki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

FEATURES  
source

1. .704  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="R0C0046N22"  
/cell\_type="1 cell"  
/dev\_stage="1 cell embryo"  
/clone\_lib="RIKEN full-length enriched, 1 cell embryo"

## ORIGIN

Query Match 81.8%; Score 687.2; DB 13; Length 704;  
Best Local Similarity 99.3%; Pred. No. 1.8e-100;  
Matches 700; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 101 ACAGTTTACCATCTGGGTGGTCTGCTGGCACTGGAGGGGGAATCAGCTTGAAGCATGG 160  
DB 1 ACAGTTTACCATCTGGG-TGGGTACGGGCACTGGAGGGGGAATCAGCTTGAAGCATGG 59  
QY 161 CAATGAATCTACATGCTCCCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGACAT 220  
DB 60 CAATGAATCTACATGCTCCCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGACAT 119  
QY 221 GTTTGTGTGACATTAATGAGCAGGACATAAGCGGGCTCCAGCATCTAAGAAGCTGAA 280  
DB 120 GTTTGTGTGACATTAATGAGCAGGACATAAGCGGGCTCCAGCATCTAAGAAGCTGAA 179  
QY 281 AAAAGCCAGTGCACCTCTTTTCATGATGCTTATACATGAGGAGCTGGCCAGT 340  
DB 180 AAAAGCCAGTGCACCTCTTTTCATGATGCTTATACATGAGGAGCTGGCCAGT 239  
QY 341 GATTTCATACCATCTAAAGCTCTGTGATGGTACCTCTCTTTTCCAGGACAGAGTT 400  
DB 240 GATTTCATACCATCTAAAGCTCTGTGATGGTACCTCTCTTTTCCAGGACAGAGTT 299  
QY 401 TAAATATTACATCAAGAGATGATCAAGGAATAGGAATATGATCTCAGGAGGCTATTATTA 460

DB 300 TAAATATTACATCAAGAGATGATCAAGGAATAGGAATATGATCTCAGGAGGCTATTATTA 359  
QY 461 CAGATAGCATGATATGTTAGTGGTACCTATTTATTGAGAACACTCTCTGAAGAGAAGGATCT 520  
DB 360 CAGATAGCATGATATGTTAGTGGTACCTATTTATTGAGAACACTCTCTGAAGAGAAGGATCT 419  
QY 521 CAAAGAAAGGATGGCTCATGCCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCGG 580  
DB 420 CAAAGAAAGGATGGCTCATGCCATGAACGAGTACCCAGACTCCTGTGGGTTCTTGTCGG 479  
QY 591 GCCTCATGGGTGTACGTGTGGGAGAAACATGGGAGAGCAAAACCAATCTGTGAGTG 640  
DB 480 GCCTCATGGGTGTACGTGTGGGAGAAACATGGGAGAGCAAAACCAATCTGTGAGTG 539  
QY 641 TTATGACTTACCTGTTTGACATTTGCTCTCCATGAAGAGAGTGGACTCGATCCAAACACA 700  
DB 540 TTATGACTTACCTGTTTGACATTTGCTCTCCATGAAGAGAGTGGACTCGATCCAAACACA 599  
QY 701 GCTCCAGTTGGAGAAATGGAATTTGTGTAAGCAAGTGGATGCTAAGCATCTCCAAACA 760  
DB 600 GCTCCAGTTGGAGAAATGGAATTTGTGTAAGCAAGTGGATGCTAAGCATCTCCAAACA 659  
QY 761 ATAAACAAACTCAATATGCTTAAATAAATCAACTCAGCTGCTTTT 805  
DB 660 ATAAACAAACTCAATATGCTTAAATAAATCAACTCAGCTGCTTTT 704

## RESULT 4

CB595118  
LOCUS  
DEFINITION  
AGENCOURT\_12932597 NIH\_MGC\_178 Mus musculus cDNA clone  
IMAGE:30301225 5', mRNA sequence.

## ACCESSION

CB595118

## VERSION

CB595118.1 GI:29512974

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 739)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Dr. Michael Brownstein

cDNA Library Preparation: Michael Brownstein Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDCM68 row: d column: 02

High quality sequence stop: 609.

Location/Qualifiers

1. .739

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:30301225"

/lab\_host="DH10B (TI-phage-resistant)"

/clone\_lib="NIH\_MGC\_178"

/note="Organ: lung and heart; Vector: pDNR-LIB; Site: 1;

sfll (ggccattatggcc); Site 2: sfll (ggccgcttggcc); cDNA

made by oligo-dT priming and directionally cloned. 5' and

3' adaptor were used in cloning as follows:

5'-AAGCAGTGTATCAACGAGTGGCATTACGGCCGG-3' and

5'-ATTTCAGAGCGGCGGCGGACATG-DT(30)NN-3'. Full-length

enriched library was constructed using the Clontech

Creator SMART kit and size-selected to contain the 0.5 kb

size fraction. Library created in the laboratory of M.

Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."

## FEATURES

source

## ORIGIN

Query Match 80.3%; Score 674.2; DB 14; Length 739;  
Best Local Similarity 99.3%; Pred. No. 2.1e-98;  
Matches 698; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
1 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 60  
2 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 60  
3 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 97  
4 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 97  
5 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 120  
6 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 120  
7 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 157  
8 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 157  
9 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 180  
10 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 180  
11 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 217  
12 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 217  
13 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 240  
14 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 240  
15 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 277  
16 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 277  
17 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 300  
18 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 300  
19 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 337  
20 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 337  
21 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 360  
22 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 360  
23 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 397  
24 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 397  
25 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 420  
26 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 420  
27 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 457  
28 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 457  
29 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 480  
30 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 480  
31 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 517  
32 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 517  
33 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 540  
34 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 540  
35 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 577  
36 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 577  
37 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 600  
38 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 600  
39 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 637  
40 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 637  
41 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 660  
42 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 660  
43 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 696  
44 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 696  
45 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 702  
46 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 702  
47 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 739  
48 GAACCATGCTGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 739

RESULT 5  
BY755126  
LOCUS  
DEFINITION  
SCA-9 clone 15 cdna Mus musculus cDNA clone G430010F18 3', mRNA  
sequence.  
ACCESSION  
BY755126  
VERSION  
BY755126.1  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 677)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.A., Maglott, D.R.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Malcais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Wagner, L., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E., and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
22354683  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
FEATURES  
source  
Location/Qualifiers  
1. 677  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Swiss Webster"  
/db\_xref="taxon:10090"  
/clone="G430010F18"  
/sex="male"  
/tissue type="submandibular gland"  
/cell\_line="CRL-1734 SCA-9 clone 15"

```

/clone lib="RIKEN full-length enriched, submandibular
gland CRL-1734 SCA-9 clone 15 cDNA"

ORIGIN
Query Match 78.7%; Score 661.2; DB 13; Length 677;
Best Local Similarity 98.7%; Pred. No. 2.7e-96;
Matches 666; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 136 GAGGGGGAATCAGCTGAAGCATGGCAATGAATCTACATCTGCTCCCTCAGCGGTGCAAA 195
DB 2 GAGCGGTAATCAGCTGAAGCATGGCAATGAATNTACATTTGCTCCCTCAGCGGTGCAAA 61
QY 196 AGAGGCGCAATCAGCCAGAGACATCTTTGTGTGTGACATTAATGAGCAGGACATAAGCG 255
DB 62 GGAAGCGCATTGAGCAGAGACATCTTTGTGTGTGACATTAATGAGCAGGACATAAGCG 121
QY 256 GGCCTTCAGCATCTAAGAGCTGAAAAAGGACAGTGCCTCTCTTTTCATGAATGCTT 315
DB 122 GGCCTTCAGCATCTAAGAGCTGAAAAAGGACAGTGCCTCTCTTTTCATGAATGCTT 181
QY 316 ATACCATGAGAGAGCTGGCGAGTGATTCATACCCACTCTAAAGCTGTGTGATGGCTA 375
DB 182 ATACCATGAGAGAGCTGGCGAGTGATTCATACCCACTCTAAAGCTGTGTGATGGCTA 241
QY 376 CCCTTCCTGTTTCAGGACAGGAGTCTTAAATTAACATCAAGAGATGATCAAGGAATAA 435
DB 242 CCCTTCCTGTTTCAGGACAGGAGTCTTAAATTAACATCAAGAGATGATCAAGGAATAA 301
QY 436 GGAATGTACCTCAGAGGCTATTACAGATACATGATGTTAGTGGTACCTATTATTG 495
DB 302 GGAATGTACCTCAGAGGCTATTACAGATACATGATGTTAGTGGTACCTATTATTG 361
QY 496 AGAACACTCCTGAAGAGAGGATCTCAAGAAAGGATGCTCATGCTCAATGAATGATACC 555
DB 362 AGAACACTCCTGAAGAGAGGATCTCAAGAAAGGATGCTCATGCTCAATGAATGATACC 421
QY 556 CAGACTCTGTGCGGTCTTGTCCGCGGTGATGCGGTGTACGTGTGGGAGAAACATGGG 615
DB 422 CAGACTCTGTGCGGTCTTGTCCGCGGTGATGCGGTGTACGTGTGGGAGAAACATGGG 481
QY 616 AGAACCAAAACCATGTGTGAGTGTATGATCACTGTTTGCATGCTGTCTCCATGA 675
DB 482 AGAACCAAAACCATGTGTGAGTGTATGATCACTGTTTGCATGCTGTCTCCATGA 541
QY 676 AGAAGATGGATCGATCAACACACCTCCAGTGTGGAGAAATGAATGTGTAGGCCA 735
DB 542 AGAAGATGGATCGATCAACACACCTCCAGTGTGGAGAAATGAATGTGTAGGCCA 601
QY 736 AGTGGATGCTAAGCATCTCCCAACATAAAACAACTCAATTATGCTTTAAATAAATCTC 795
DB 602 AGTGGATGCTAAGCATCTCCCAACATAAAACAACTCAATTATGCTTTAAATAAATCTC 661
QY 796 AGCTGCTTTTAAAAA 810
DB 662 ATCTGCTTTTAAAAA 676

RESULT 6
AV306484 LOCUS
AV306484 RIKEN full-length enriched, 8 days embryo Mus musculus
DEFINITION cDNA clone 5730540M11 3' similar to AB028863 Mus musculus mRNA for
WDRP19, mRNA sequence.
ACCESSION AV306484
VERSION AV306484.2 GI:16393817
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 645)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,

```

Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Nov 10, 1999 this sequence version replaced gi:6338998.

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamakawa,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers  
1. .645  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="5730540M11"  
/sex="mixed"  
/dev\_stage="8 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 8 days embryo"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATCTCGAGTCTATTAAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(-) after bulk excision from Lambda



FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

```
ORIGIN
Query Match      76.4%; Score 641.8; DB 9; Length 645;
Best Local Similarity 99.7%; Pred. No. 3.4e-93;
Matches 643; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 140 GCGAATCAGCTTGAAGCATGCGCAATGAATCTACATTGCTCCCTCAGGCGGTGCAAAAGGA 199
DB 1 GGGATCAGCTTGAAGCATGCGCAATGAATCTACATTGCTCCCTCAGGCGGTGCAAAAGGA 60
QY 200 GCGAATCAGCTTGAAGCATGCGCAATGAATCTACATTGCTCCCTCAGGCGGTGCAAAAGGA 259
DB 61 ACGAATCAGCTTGAAGCATGCGCAATGAATCTACATTGCTCCCTCAGGCGGTGCAAAAGGA 120
QY 260 TCCAGCATCTAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAG 319
DB 121 TCCAGCATCTAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAGAGCTGAAG 180
QY 320 CATGAGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAGCTGCTGTGATGGCTACCCCT 379
DB 181 CATGAGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAGCTGCTGTGATGGCTACCCCT 240
QY 380 TCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGAGATCAAGAGATCAAGAGAA 439
DB 241 TCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGAGATCAAGAGATCAAGAGAA 300
QY 440 ATGTACCTCAGGAGGCTTATACAGATGATGATGATGATGATGATGATGATGATGATGAT 499
DB 301 ATGTACCTCAGGAGGCTTATACAGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 500 CACTCTCTGAAGAGAGGATCTCAAGAGAGAGATGGCTCATGCCATGATGATGATGATGATG 559
DB 361 CACTCTCTGAAGAGAGGATCTCAAGAGAGAGATGGCTCATGCCATGATGATGATGATGATG 420
QY 560 CTCCTGTGCGGTTCTTGTGCGGCTCATGAGGCTTATGATGATGATGATGATGATGATGATG 619
DB 421 CTCCTGTGCGGTTCTTGTGCGGCTCATGAGGCTTATGATGATGATGATGATGATGATGATG 480
QY 620 AGCAAAACCATGCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 679
DB 481 AGCAAAACCATGCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 680 GATGGGACTCGATCCACACAGCTCCAGTGGAGAGAAATGGAATGTTGTAAGCAAGTG 739
DB 541 GATGGGACTCGATCCACACAGCTCCAGTGGAGAGAAATGGAATGTTGTAAGCAAGTG 600
QY 740 GATGGCTAAGCATCTCAACAAATTAACAACTCAATATATGCTT 784
DB 601 GATGGCTAAGCATCTCAACAAATTAACAACTCAATATATGCTT 645
```

```
RESULT 7
BG277254
LOCUS
DEFINITION
  BG277254 679 bp mRNA linear EST 21-FEB-2001
  ux41d07.y1 Soares.NMAX maxillary process Mus musculus cDNA clone
  IMAGE:3512869 5' similar to RF:Q9WV05 Q9WV05 MMRP19. ;, mRNA
  sequence.
ACCESSION
  BG277254
VERSION
  BG277254.1 GI:13072373
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 679)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Other ESTs: ux41d07.x1
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
```

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1393445  
Seq primer: -40RP from Gibco  
High quality sequence stop: 475.  
Location/Qualifiers  
1. .679

#### FEATURES

source  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3512869"  
/tissue\_type="maxillary process"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Soares.NMAX maxillary process"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA  
was primed with a NotI - oligo(dT) primer [5',  
TGTTCAATCTGAAGTGGAGGCGGCGGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with NotI and cloned into the NotI  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

#### ORIGIN

```
Query Match      73.8%; Score 620; DB 12; Length 679;
Best Local Similarity 99.2%; Pred. No. 1e-89;
Matches 623; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAACCATGTCTGCTGTCAAGCTCAAGGAGACTGTGCTCGCGCGCTGTGGCGCGAGG 60
DB 48 GAACCATGTCTGCTGTCAAGCTCAAGGAGACTGTGCTCGCGCGCTGTGGCGCGAGG 107
QY 61 ACAGGAGCACCCCGATCTCCATCCAGAACTTTGCAAAACAGTTTACCATCTGGCT 120
DB 108 ACAGGAGCACCCCGATCTCCATCCAGAACTTTGCAAAACAGTTTACCATCTGGCT 167
QY 121 GGGTCACTGTCAGCTGAGGCGGGAATCAGTTGGAAGCATGGCAATGAATCTACATTGCTC 180
DB 168 GGGTCACTGTCAGCTGAGGCGGGAATCAGTTGGAAGCATGGCAATGAATCTACATTGCTC 227
QY 181 CTCAGCGCTGCAAAAGAGAGCGCATTCAGCAGAGACATGTTTGTGTGACATTAATG 240
DB 228 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCAGAGACATGTTTGTGTGACATTAATG 287
QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCACTCTC 300
DB 288 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCACTCTC 347
QY 301 TTTTTCATGAATGCTTATACCATGAGAGAGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360
DB 348 TTTTTCATGAATGCTTATACCATGAGAGAGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 407
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTAACATCAAGAGA 420
DB 408 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTAACATCAAGAGA 467
QY 421 TGATCAAGGAATAAGGAATGATACCTCAGAGGCTATTACAGATAGATGATGTTAG 480
DB 468 TGATCAAGGAATAAGGAATGATACCTCAGAGGCTATTACAGATAGATGATGTTAG 527
QY 481 TGATCAAGGAATAAGGAATGATACCTCAGAGGAGAGATCTCAAGAGAGGATGCTCATG 540
DB 528 TGATCAAGGAATAAGGAATGATACCTCAGAGGAGAGATCTCAAGAGAGGATGCTCATG 587
QY 541 CCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 588 CCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 647
QY 601 GGGGAGGAACATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 628
DB 648 GGGGAGGAACATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
```

```

RESULT 8
AW988838
LOCUS      AW988838               642 bp      mRNA      linear      EST 02-JUN-2000
DEFINITION uf21b06.y1 Soares mammary_gland NLMG Mus musculus cDNA clone
IMAGE:1511987 5' similar to SW:YJZ4 YEAST P47095 HYPOTHETICAL 27.4
KD PROTEIN IN MER2-BNA1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION  AW988838
VERSION     AW988838.1  GI:8184083
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 642)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:938839
Seq primer: -40RP from Gibco
High quality sequence stop: 469.
FEATURES             source
     1..642
     /organism="Mus musculus"
     /mol_type="mRNA"
     /db_xref="taxon:10090"
     /clone="IMAGE:1511987"
     /sex="female (lactating)"
     /tissue_type="mammary gland"
     /lab_host="DH10B"
     /clone_lib="Soares mammary gland NLMG"
     /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaudo."
ORIGIN
Query Match      73.6%; Score 618.6; DB 10; Length 642;
Best Local Similarity 99.2%; Pred. No. 1.7e-89;
Matches 621; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  GAACCATGTCGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60
DB      17  GAACCATGTCGGCTGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 76

QY      61  ACAAGGAGCACCCTGATTCCTGATCCCAAGACTTTTGCAACAGTTTACCACTGGGCT 120
DB      77  ACAAGGAGCACCCTGATTCCTGATCCCAAGACTTTTGCAACAGTTTACCACTGGGCT 136

QY      121  GGGTCACTGGGCTGAGGGGGAATCAGTTGAAGCATGGCAATGAATCTACATTGCTC 180
DB      137  GGGTCACTGGGCTGAGGGGGAATCAGTTGAAGCATGGCAATGAATCTACATTGCTC 196

QY      181  CCTCAGCGCTGCAAGAGGAGCGCATTCAGCAGACACATGTTGTGTGACATTAAATG 240
DB      197  CCTCAGCGCTGCAAGAGGAGCGCATTCAGCAGACACATGTTGTGTGACATTAAATG 256

QY      241  AGCAGGACATAGCGGGCTCCAGCATCTTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 300
DB      257  AGCAGGACATAGCGGGCTCCAGCATCTTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 316

QY      301  TTTTCAATGAATGTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 360
DB      317  TTTTCAATGAATGTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 376

```

```

QY      361  CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
DB      377  CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 436

QY      421  TGATCAAGGAATTAAGGAATGATCTCCTCAGGAGCTTATACAGATACGATGATGTTAG 480
DB      437  TGATCAAGGAATTAAGGAATGATCTCCTCAGGAGCTTATACAGATACGATGATGTTAG 496

QY      481  TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGGATGCTCATG 540
DB      497  TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGGATGCTCATG 556

QY      541  CCATGAATGATGATCCAGACTCTCTGCGGTTCTGTCCGCGCTCATGGGTTGACGTGT 600
DB      557  CCATGAACGAGTACCCAGACTCTCTGCGGTTCTGTCCGCGCTCATGGGTTGACGTGT 616

QY      601  GGGGAGAAACATGGAGAAAGCAAAA 626
DB      617  GGGGAGAAACATGGAGAAAGCAAAA 642

RESULT 9
BU558424
LOCUS      BU558424               850 bp      mRNA      linear      EST 16-SEP-2002
DEFINITION AGENCOURT 1036389 NIH MGC 144 Mus musculus cDNA clone
IMAGE:6586918 5', mRNA sequence.
ACCESSION  BU558424
VERSION     BU558424.1  GI:22908720
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 850)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2801 row: d column: 22
High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     1..850
     /organism="Mus musculus"
     /mol_type="mRNA"
     /db_xref="taxon:10090"
     /clone="IMAGE:6586918"
     /lab_host="DH10B (TI-phage-resistant)"
     /clone_lib="NIH MGC 144"
     /note="Organ: Brain; Vector: pDNR-LIB; Site 1: SfiI
(ggcatatggcc); Site 2: SfiI (ggcgcctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGATGGCCATTACGGCCGG-3' and
5'-ATTCTAGAGCGGAGCGGCGGACATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.2-0.5
kb size fraction (other fractions present in NIH MGC 143).
Library created in the laboratory of M. Brownstein (NIMH,
NIH). Note: this is a NIH_MGC Library."
ORIGIN
Query Match      72.1%; Score 605.6; DB 13; Length 850;
Best Local Similarity 91.4%; Pred. No. 1.8e-87;
Matches 729; Conservative 0; Mismatches 14; Indels 55; Gaps 6;

```

JOURNAL  
COMMENT

Unpublished (2001)  
On Feb 13, 2001 this sequence version replaced gi:12789030.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5727.r For  
more information about this cluster, see

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC012BF01Q1&cluster=5727.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DC012BF01Q1.

FEATURES  
source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC012YK02"  
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 71.1%; Score 597; DB 9; Length 1201;  
Best Local Similarity 84.8%; Pred. No. 3.7e-86;  
Matches 693; Conservative 0; Mismatches 120; Indels 4; Gaps 2;  
4 CCATGCTGCTGCTCAAGCTC---AAGGAGACTGTGCTCGGGCCGTGTGGCGGCGAGG 60  
Db CCATGCTGCTGCTGCTGCTCGGGAGGAGACTGTGCTCCGGAGATGGCGGCGAGG 155  
61 ACAAGGAGCAGCCCGGATCTCTGATCCAGACTTTTGCAGACAGTTTACCATCTGGCT 120  
Db ACAAGGAGCATCCAGATACCTGATCCAGACTTTTGCAGACAGTTTACCATCTGGCT 215  
121 GGCTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTC 180  
Db GGCTCACTGGCACTGGAGGAGGAATAGCTTGAAGCATGGGATGAATCTACATTGCTC 275  
181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240  
Db CTTCAAGAGTGCAAAAGGAGCGAATTCAGCTGAGACATGTTTGTGTGATATAATG 335  
241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAGAGCGAAGAGCGAGTCTCTC 300  
Db AAAAGGACATAAGTGAGACCTTCGCCATCGAAGAGCTTAAAGAGCGAGTCTCTCTC 395  
301 TTTTCATGATGCTTATACCATGAGAGGAGCTGGCGAGTGTTCATACCCACTCTAAAG 360  
Db TTTTCATGATGCTTATACCATGAGAGGAGCTGGCGAGTGTTCATACCCACTCTAAAG 455  
361 CTGCTGTGATGCTTACCCCTTCTGTTTCCAGACAGGAGTGTTPAAATTAACATCAAGAGA 420  
Db CTGCTGTGATGCTTACCCCTTCTGTTTCCAGACAGGAGTGTTPAAATTAACATCAAGAGA 515  
421 TGATCAAGGAATTAAGGAATGTGCTCCTCAGAGGCTTATACAGATACGATGATGTTAG 480  
Db TGATCAAGGAATTAAGGAATGTGCTCCTCAGAGGAGTGTTPAAATTAACATCAAGAGA 575  
481 TGGTACCTTATTTGAGNACACTCTGAGAGAGAGGATCTCAAGAGAGGAGTGGCTCATG 540  
Db TGGTACCTTATTTGAGNACACTCTGAGAGAGAGGATCTCAAGAGAGGAGTGGCTCATG 635  
541 CCATGAATGAGTACCAGACTCTGCTGGGTTCTTGTCCGGCGTTCATCGGGTGTACGTGT 600  
Db CAATGAATGAGTACCAGACTCTGCTGGGTTCTTGTCCGGCGTTCATCGGGTGTACGTGT 695  
601 GGGGAGAAACATCGGAGAGAAACCAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660

1 GAACCATGTGCTGCTCAAGCTCAAGGAGACTGTGCTCGGGCGGTGTGGCGGCGAG 60  
Db GAACCATGTGCTGCTGCTCAAGCTCAAGGAGACTGTGCTCGGGCGGTGTGGCGGCGAG 151  
61 ACNAGGAGCAGCCCGGATCTCTGATCCAGAACTTTGCAAAACAGTTTTTACCATCTGGCT 120  
Db ACNAGGAGCAGCCCGGATCTCTGATCCAGAACTTTGCAAAACAGTTTTTACCATCTGGCT 211  
121 GGCTCACTGCTGCTGAGGGGGAATCAGCTTGAAGCATGGAAGTGAATTAATCAATTGCTC 180  
Db GGCTCACTGCTGCTGAGGGGGAATCAGCTTGAAGCATGGAAGTGAATTAATCAATTGCTC 250  
181 CCTCAGCGCTGCAAAAGGAGCGCAATTCAGCCAGAGACATGTTGTTGTTGACATTAATG 240  
Db CCTCAGCGCTGCAAAAGGAGCGCAATTCAGCCAGAGACATGTTGTTGTTGACATTAATG 282  
241 AGCAGGACATAAGCGGCTCCAGATCTAAGAGCTGMAAAAGCCAGTGCACCTCTC 300  
Db AGCAGGACATAAGCGGCTCCAGATCTAAGAGCTGMAAAAGCCAGTGCACCTCTC 342  
301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGTGTATCAATCCACTCTAAAG 360  
Db TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGTGTATCAATCCACTCTAAAG 402  
361 CTGCTGTGATGCTGCTGCTTCCAGACAGGAGTGTAAATTAACATCAACATCAAGAGA 420  
Db CTGCTGTGATGCTGCTGCTTCCAGACAGGAGTGTAAATTAACATCAACATCAAGAGA 462  
421 TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTTATACAGATACGATGATGTTAG 480  
Db TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTTATACAGATACGATGATGTTAG 522  
481 TGGTACCTTATTTGAGACACCTCTGAGAGAGAGATCTCAAGAGAGATGCTCATG 540  
Db TGGTACCTTATTTGAGACACCTCTGAGAGAGAGATCTCAAGAGAGATGCTCATG 582  
541 CCATGAATGAGTACCCAGACTCTGCTGGGTTCTTGTCCGGCGTTCATGGGTTGACGTGT 600  
Db CCATGAATGAGTACCCAGACTCTGCTGGGTTCTTGTCCGGCGTTCATGGGTTGACGTGT 642  
601 GGGGAGAAACATGGAG-AAAAGAAAACCATGTGTGAGTGTATGACTACCTGTTTGAC 659  
Db GGGGAGAAACATGGAGAAAAGAAAACCATGTGTGAGTGTATGACTACCTGTTTGAC 702  
643 GGGGAGAAACATGGAGAAAAGAAAACCATGTGTGAGTGTATGACTACCTGTTTGAC 716  
660 ATTGCTGTCTCCATG-AAGAGATGGGACTCGATCCACACAGCTTCCA--GTTGGAGAA 762  
703 ATTGCTGTCTCCATGAAAAGATGGGACTCGATCCACACAGCTTCCAAGATGGAGAAA 774  
717 AATGAATGTTGT-AAGCAAGTGATG-CCTAAGCATCTCCAACTAATAAACAACCTCA 774  
763 AATGAATGTTGTAAAGCCAGTGTGATGCCCTTAAGCATCTCCAACTAATAAACAACCTCC 822  
775 ATTATGCTTAAATAAAA 792  
823 ATTATGCTTAAATAAAA 840

RESULT 10  
AL525537 1201 bp mRNA linear EST 23-MAY-2003  
LOCUS  
DEFINITION  
CDNA clone CS0DC012YK02 5-PRIME, mRNA sequence.  
ACCESSION  
AL525537.2 GI:31063401  
VERSION  
AL525537  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization

```

696 GGGGGAACATGGGAGAGCCAAACCATGTGTGAGTGTATGACTATTATTATGATA 755
661 TTGCTGTCTCCATGAAGAAGATGGGAGCTCGATCCAAACAGAGCTCCAGTGTGAGAAATG 720
756 TTGCGGTATCAATGAAGAAGTAGGACTTGTATCTTCCAGAGTCCAGTGTGAGAAATG 815
721 GAATTTGTGAAGCCAGTGGATGCTTAAGCATCTCCAAACAATAAACA-AAAATCAATTAT 779
816 GAATTTGTGAAGCCAAAGAAAGTCTTAATATATACAGATTAAGCTAAACGTAATTAT 875
780 GCCTTAATAAATAACTCAGCTGCTCTTTTAAAAAAGAAAA 816
876 TATTTAATGAAGTCTATTTTAAATGAATGAAA 912

BX362087 1034 bp mRNA linear EST 05-MAY-2003
BX362087 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ010YP07 5-PRIME, mRNA sequence.
BX362087
BX362087.1 GI:30380601
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1034)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5727.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ010CH04QPLcluster=5727.r. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DJ010CH04QPL.
Location/Qualifiers
1..1034
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ010YP07"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source
1..1034
Query Match 71.0%; Score 596.4; DB 13; Length 1034;
Best Local Similarity 84.0%; Pred. No. 4.9e-86;
Matches 687; Conservative 8; Mismatches 119; Indels 4; Gaps 2;

QY 4 CCATGTCTGGCTGTCAAGCTC---AAGGAGACTGTCTCGCGGCGGTGTGGCGGCGAGG 60
DB 94 CCATGTCTGGCTGTGTGATGCTCGGAGGAGAGACTGTCTCCCGAGATGCGGCGCGAGG 153
QY 61 ACAAGAGACACCCCGGATCTCTGATCCAGAACTTTCGAAACAGTTTTTACCATCTGGCT 120
DB 154 ACAAGAGACATCCAGATACCTGTATCCAGAACTTTCGAAACAGTTTTTACCATCTAGGCT 213
QY 121 GGGTCACTGGCACTGGAGGGGAGATCAGCTTGAAGCATGGCAATGAAATCTACATTCCTC 180
```

```

214 GGGTCACTGGGCACTGGAGAGGAATTAGCTTGAAGCATGGCGATGAAATCTACATTCCTC 273
181 COTCAGGCGCTGCAAAAGGAGCGCATTCAGCCAGAAAGACATGTTTGTGTGTGACATTAATG 240
274 CTTTCAAGGAGTGCATAAGGAGCAATTCAGCCTTGAAGACATGTTTGTGTGTGATATAATG 333
241 AGCAGACATAAAGCCGGCTCCAGCATCTAAGAAGCTGAATAAAGCCAGTGCACCTCCCTC 300
334 AAAAGCACAATAAGTGGACCTTCGCCATCGAAGAGCTAAAAAAGCCAGTGTACTCCCTC 393
301 TTTTCATGAATCTTATACCATGAGAGAGCTGGGCGAGTGTATCATACCCCTCTTAAG 360
394 TTTTCATGAATCTTATACCATGAGAGAGCTGGGCGAGTGTATCATACCCCTCTTAAG 453
361 CTGCTGTGATGGCTACCCCTTCTTCCAGGACAGAGTGTAAATTAACATCAAGAGA 420
454 CTGCTGTGATGGCCACACTTCTTCCAGGACGCGGAGTTTAAATTAACATCAAGAGA 513
421 TGTATCAAGAAATAAGAAATGTACCTCAGGAGGCTTATACAGATACGATGATGTATG 480
514 TGATAAAGGAAWAAAGAAATGTACTTCCGAGGGTATTATAGATATGATGATGTATG 573
481 TGGTACTCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAAGAAAGATGGCTCATG 540
574 TGTATCCCATTTATTGAGATACACTTGAGGAGAAARACCTCAAAATAGATAATGGCTCATG 633
541 CCATGAATAGTACCAGACTCTCTGTGCGGTTCTTGTCCGCGCTCATGGGGTGTACGTGT 600
634 CAATGAATGAATACCCAGACTCTCTGTGCGAGTCTGTGTCTCARACGTCATGGAGTATATGTGT 693
601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACTCTGTTTGACA 660
694 GGGGGGAAACATGGGAGAGGCCAAACCATGTGTGAGTGTATGACTACTCTGTTTGACA 753
661 TTGCTGTCTCCATGAAGAGATGGGACTCCAGTCCAAACACAGCTCCCGAGTGGAGAAATG 720
754 TTGCGGTATCAAGAAGAAATAGGACTTGATCTTCCAGCTCCAGTGGAGAAATG 813
721 GAATTTGTGAAGCAAGTGGATGCTTAAGCATCTCCAAACATTAAC-AAAATCAATTAT 779
814 GAATTTGTGAAGCAAGAAAGTCTTAATATATACAGATTAAGCTTAAAGTAAATAT 873
780 GCCTTAATAAATCAGCTGCTCTTTTAAAAAAGAAAA 817
874 TATTTAATGAAGCTATTTTAAATGAATGAAW 911

RESULT 12
BQ193148/6
LOCUS
DEFINITION
BQ193148
ACCESSION
BQ193148.1 GI:20368699
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 645)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA
Tel: 319 335 8250
```

Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA Library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 1-22, >AT-rich#low\_complexity  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

source

Location/Qualifiers

1..645  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-DRI-clk-f-14-0-UI"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-DRI"  
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in p317 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery, Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AAGATATCAA between the Not I cloning site and d18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School.  
 TAG TISSUE=osteoblast  
 TAG LIB=UI-R-DRI  
 TAG\_SEQ=AAGATATCAA"

## ORIGIN

Query Match 69.7%; Score 585.8; DB 13; Length 645;  
 Best Local Similarity 94.3%; Pred. No. 3e-84;  
 Matches 608; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 179 TCCTCAGCGGTGAAAGAGAGCGCATTCAGCAGAGACATGTTGTTGTGACATTA 238  
 Db 645 TCCCTCAGCGGTGAAAGAGAGCGCATTCAGCAGAGACATGTTGTTGTGACATTA 586

QY 239 TGAGCAGGACATAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCATCC 298  
 Db 585 TGAGCAGGACATAGTGGGCTCCAGCATCTAAGAGCTTAAAGAGCCAGTGCATCC 526

QY 299 TCTTTTCATGAATGCTTATACCATGAGAGGAGTGGCGGAGTGATTCATCCACTTAA 358  
 Db 525 TCTTTTCATGAACGCTTACACCATGAGAGGAGCGGCGAGTGATTCATCCACTTAA 466

QY 359 AGTCGCTGTGATGCTACCCCTCTGTTTCCAGGACAGAGTTAAATTAACATCAAGA 418  
 Db 465 AGCTGCTGTGATGCTACCCCTCTGTTTCCAGGACAGAGTTAAATTAACATCAAGA 406

QY 419 GATCATCAAGGATAGGAATGTACCTCAGGAGGCTATTACAGATCAGATGATATGTT 478  
 Db 405 GATCATCAAGGATAGGAATGTACCTCTGGAGGGTATTACAGATCAGATGATATTT 346

QY 479 AGTGATCACTTATTGAGAACACTCTCTGAAGAGAGAGGATCTCAAGAAAGATGGCTCA 538  
 Db 345 AGTGATCACTTATTGAGAACACTCTCTGAAGAGAGAGGACCTCAAGAAAGATGGCTCG 286

QY 539 TGCCATGAATGAGTACCCAGACTCTGTGCGGTCTTTGTCGGCGCTCATGGGTGTAGCT 598  
 Db 285 TGCCATGAATGAATACCCAGACTCTGTGCGGTCTTTGTCAGGGCTCAGGGGTGTAGCT 226

QY 599 GTGGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGA 658

Db 225 GTGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGATTTACTGTTGA 166  
 QY 659 CATTGCTCTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAA 718  
 Db 165 CATTGCTCTCTCCATGAAGAAGATGGGCTGGATCCACGCGAGTTCCTCCAGTTGGAGACA 106  
 QY 719 TGGAAATTTGTAAGCAAGTGGATGCCCTTAAGCATCTCCAAACATTAACAACACTCAATTA 778  
 Db 105 TGGAAATCGTATAAGCAAGTGGAGCGCTTAAGAACTCTCCAAATAATAACAACACTCAATTA 46  
 QY 779 TGCCTTAATAAAGCTCAGCTCTTTTAAAAAATAAAAAAATAAAAAA 823  
 Db 45 CGCCTTAATAAAGCTCAGCTCTTTTAAAAAATAAAAAAATAAAAAA 1

## RESULT 13

AL559009

LOCUS

DEFINITION

Homo sapiens cDNA clone CS0DJ008YC08 5-PRIME, mRNA sequence.

ACCESSION

AL559009.2

VERSION

EST.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 1124)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 15, 2001 this sequence version replaced gi:12904084.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by life technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

5727.r For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DJ008BB04QP1&amp;cluster=5727.r. Contact :

Feng Liang Email : fliang@lifetech.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0DJ008BB04QP1.

Location/Qualifiers

1..1124

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DJ008YC08"

/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"

/cell\_line="JURKAT"

/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 69.6%; Score 585; DB 9; Length 1124;  
 Best Local Similarity 84.5%; Pred. No. 3.2e-84;  
 Matches 692; Conservative 1; Mismatches 121; Indels 5; Gaps 3;

QY 4 CCATGTCTGCTGCAAGCTC---AAGGAGACTGTTCTCGCGCCCTGTGGCGCGCAGG 60  
 Db 138 CCATGTCTGCTGCTGCTGCTGCGAGGAGACTGTTGTTCCCGAGAGTGGCGCGCAGG 197

QY 61 ACAGGAGGACCCCGGATTCCTGATCCAGAACTTTGCAAAACAGTTTTTACCATCTGGCT 120  
 Db 198 ACAGGAGGATCCAGATACCTGATCCAGNACTTTGCAACAGTTTTTACCATTAGGCT 257

QY 121 GGGTCACTGGCACTGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180  
 Db 258 GGGTCACTGGCACTGAGGGGGAATGAGCTTGAAGCATGGCAATGAAATCTACATGCTC 317  
 QY 181 CTTCAAGCGCTGCAAAAGAGGCGCATTCAGCCAGACACATGTTTGTGTGATTAATG 240  
 Db 318 CTTCAAGCGCTGCAAAAGAGGCGCATTCAGCCAGACACATGTTTGTGTGATTAATG 377  
 QY 241 AGCAGACATAGCGGGCTCCAGCATCTAAGAGCT-GAAAAAGCGAGTGCCTCT 299  
 Db 378 AAAAGACATAGCGGGCTCCAGCATTCAGCCAGACACATGTTTGTGTGATTAATG 437  
 QY 300 CTTTTCATGAATGCTTATACCATGAGAGGAGTGGCGGAGTGAATCATCCCACTTAA 359  
 Db 438 CTTTTCATGAATGCTTATACCATGAGAGGAGTGGCGGAGTGAATCATCCCACTTAA 497  
 QY 360 GCTGCTGTGATGCTACCTCTCTGTTTCCAGGACAGAGTGAATTAACATCAAGAG 419  
 Db 498 GCTGCTGTGATGCTACCTCTCTGTTTCCAGGACAGAGTGAATTAACATCAAGAG 557  
 QY 420 ATGATCAAGGATTAAGGAATGCTTCTGAGGAGTGAATTAACATCAAGAG 479  
 Db 558 ATGATCAAGGATTAAGGAATGCTTCTGAGGAGTGAATTAACATCAAGAG 617  
 QY 480 GTGCTACCTATTATGAGACACTCTGAGAGAGGATCTCAAGAGAGGATGGCTCAT 539  
 Db 618 GTGCTACCTATTATGAGACACTCTGAGAGAGGATCTCAAGAGAGGATGGCTCAT 677  
 QY 540 GCCATGAATGAGTACCCAGACTCTGCTGCGGTTCTTGTCCGCGCTCATGGGTTACG 599  
 Db 678 GCAATGAATGAGTACCCAGACTCTGCTGCGGTTCTTGTCCGCGCTCATGGGTTACG 737  
 QY 600 TGGGGAGAACATGGGAGAGAACCAATCATGCTGAGTGTATGACTACCTGTTTAC 659  
 Db 738 TGGGGAGAACATGGGAGAGAACCAATCATGCTGAGTGTATGACTACCTGTTTAC 797  
 QY 660 ATTGCTGTCTCCATGAAGAGAGTGGGACTCGATCCCAACAGCTCCCGAGTTGAGAGAA 719  
 Db 798 ATTGCTGTCTCCATGAAGAGAGTGGGACTCGATCCCAACAGCTCCCGAGTTGAGAGAA 857  
 QY 720 GGAATTGTGTAAAGCAAGTGGTCTTAAGCATCTCCCAACAGCTCCCGAGTTGAGAGAA 778  
 Db 858 KGAATTGTCTAAGCCAAAGAGAGTCTTAATATATACAGAGATAAAGCTAAAGCTTA 917  
 QY 779 TGCCTTAATTAATCACTGAGTGTCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAG 817  
 Db 918 TTAATTAATGACAGCTATTTTCTTAATGAATTGAGAA 956

RESULT 14  
 CF621082 597 bp mRNA linear EST 02-OCT-2003  
 LOCUS laf22a02.y1 Gastric Epithelial Progenitor Mus musculus cDNA 5'  
 DEFINITION similar to TR:Q9WV05 Q9WV05 MMRP19.1, mRNA sequence.

CF621082  
 ACCESSION CF621082.1 GI:37370011  
 VERSION EST.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus

ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 597)  
 REFERENCES Tidwell,R., Clifton,S., Marra,M., Hillier,L., Page,D., Martin,J., Wyllie,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennet,J., Ronko,I., Tsagaris,I., R., Belaygorod,L., Grow,A., Maguire,L., Waterston,R. and Wilson,R.  
 TITLE WashU Stem cell EST Project  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Jeff Gordon and Mike Lovett  
 Washington University School of Medicine  
 1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified

SMART primers. The final cDNA was cloned in pAMP1 vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI.

Seq primer: -40RP from Gibco  
 High quality sequence stop: 585.  
 Location/Qualifiers

FEATURES  
 source

1..597  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="Gastric Epithelial Progenitor"  
 /dev\_stage="adult"  
 /lab\_host="DH5alpha"  
 /note="Vector: pAMP1; This library was created from laser-captured isthmal cells from tox176 transgenic mice. 1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified SMART primers. The final cDNA was cloned in pAMP1 vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI."

ORIGIN

Query Match 68.7%; Score 577; DB 14; Length 597;  
 Best Local Similarity 98.3%; Pred. No. 7.8e-83;  
 Matches 593; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 124 TCATCGCCTGAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTCCCT 183  
 Db 5 TAAACCTCTCACTAAAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTCCCT 64  
 QY 184 CAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATGAGC 243  
 Db 65 CAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATGAGC 124  
 QY 244 AGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCTTT 303  
 Db 125 AGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCTTT 184  
 QY 304 TCATGAATGTTATACCATGAGAGAGCTGGCCAGTGTTCATACCCCTCTTAAGCTG 363  
 Db 185 TCATGAATGTTATACCATGAGAGAGCTGGCCAGTGTTCATACCCCTCTTAAGCTG 244  
 QY 364 CTGTGATGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGATGA 423  
 Db 245 CTGTGATGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGATGA 304  
 QY 424 TCAAGGAATAAGGAAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAGTGG 483  
 Db 305 TCAAGGAATAAGGAAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAGTGG 364  
 QY 484 TACCTATTATGAGACACTCTCTGAGAGAGAGATCTCAAGAGAGAGTGGCTCATGCCA 543  
 Db 365 TACCTATTATGAGACACTCTCTGAGAGAGAGATCTCAAGAGAGAGTGGCTCATGCCA 424  
 QY 544 TGAATGATGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGATGA 603  
 Db 425 TGAACGATGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGATGA 484  
 QY 604 GAGAAACATGGGAG 663  
 Db 485 GAGAAACATGGGAG 544  
 QY 664 CTGTCTCCATGAGAGAGATGGGACTCGATCCCAACAGAGCTCCCAAGTTGGAGAA 716  
 Db 545 CTGTCTCCATGAGAGAGATGGGACTCGATCCCAACAGAGCTCCCAAGTTGGAGAA 597

RESULT 15  
 BY735766

LOCUS BY735766 710 bp mRNA linear EST 17-DEC-2002  
 DEFINITION cDNA clone I0C0046N22 5', mRNA sequence.

ACCESSION BY735766  
 VERSION BY735766.1 GI:27148893  
 KEYWORDS EST.

# ORGANISM

Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

# REFERENCE

1 (bases 1 to 710)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, K.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schombach, C., Gojohori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
 Ciothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
 Sanderlin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
 Verdard, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L.G., Wysshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, L., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs

# TITLE

Nature 420, 563-573 (2002)

22354683  
 12468851

# JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
 Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
 Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
 Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
 Nomura, K., Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N.,  
 Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
 Takeda, Y., Waki, K., Watanabe, K., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken

# FEATURES

source  
 1..710  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="I0C0046N22"  
 /cell\_type="1 cell"  
 /dev\_stage="1 cell embryo"  
 /clone\_lib="RIKEN full-length enriched, 1 cell embryo"

# ORIGIN

Query Match 68.4%; Score 574.6; DB 13; Length 710;  
 Best Local Similarity 98.3%; Pred. No. 1.8e-82;  
 Matches 591; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTCGCGCGCGTGTGGCGGAGG 60  
 DB 111 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTCGCGCGCGTGTGGCGGAGG 170  
 QY 61 ACAAGGAGCAGCCCGGATCTCTGATCCAGAACTTTGCAACAGCTTTTACCATTGGGCT 120  
 DB 171 ACAAGGAGCAGCCCGGATCTCTGATCCAGAACTTTGCAACAGCTTTTACCATTGGGCT 230  
 QY 121 GGGTCACCTGGCAGCTGGAGGGGGAATCAGTTGAAGCATGGCAATCAAAATCTACATTGGCTC 180  
 DB 231 GGGTCACCGGACCTGGAGGGGGAATCAGTTGAAGCATGGCAATCAAAATCTACATTGGCTC 290  
 QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAAATG 240  
 DB 291 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAAATG 350  
 QY 241 AGCAGGACATAAGCGGGCTCAGCATCTAGAGAGCTGAAAAAGCCAGTGCCTCCTC 300  
 DB 351 AGCAGGACATAAGCGGGCTCAGCATCTAGAGAGCTGAAAAAGCCAGTGCCTCCTC 410  
 QY 301 TTTTCATCAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360  
 DB 411 TTTTCATCAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 470  
 QY 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGAGTTTAAATACATCAAGAGA 420  
 DB 471 CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGAGTTTAAATACATCAAGAGA 530  
 QY 421 TGATCAAAAGGATAAGGAAATGTACTCAGGAGGCTATTACAGATACGATGATGTATG 480  
 DB 531 TGATCAAAAGGATAAGGAAATGTACTCAGGAGGCTATTACAGATACGATGATGTATG 590  
 QY 481 TGGTACCTATTATTGAGAACTCTCTGAGAGAGAGTCTCAAGAAAGGATGGCTCATG 540  
 DB 591 TGGTACCTATTATTGAGAACTCTCTGAGAGAGAGTCTCAAGAAAGGATGGCTCATG 649  
 QY 541 CCATGAATGAGTACCCAGACCTCTCTGCGGTTCTTTGTCCGCGCTCATGGGGGTGACGTG 600  
 DB 650 CCATGAACGAGTACCCAGACCTCTCTGCGGTTCTTTGTCCGCGCTCATGGGGGTGACGTG 709  
 QY 601 G 601  
 DB 710 G 710

Search completed: June 2, 2004, 01:01:34  
 Job time : 3636 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 1, 2004, 13:19:15 ; Search time 60 Seconds  
(without alignments)  
1134.899 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCRCRPGCAQDK.....SMKMKGLDPTQLPVGENGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1317	100.0	241	4	AAy85635 Antigen r
2	1317	100.0	241	5	AAU77177 Murine G-
3	1317	100.0	241	5	AAU78361 Cell diff
4	1246.5	94.6	242	4	AAg67127 Amino aci
5	1239.5	94.1	242	5	AAU78360 Cell diff
6	1235.5	93.8	242	2	AAW94762 Amino aci
7	1235.5	93.8	242	4	AAy85636 Antigen r
8	1235.5	93.8	242	5	AAU77178 Human G-C
9	786	59.7	227	4	ABb65485 Drosophil
10	248	18.8	129	4	AAo10783 Human pol
11	202	16.9	59	4	AAg74374 Human col
12	201	15.3	212	6	ABU17451 Protein e
13	195	14.8	204	6	ABU41912 Protein e
14	172	13.1	204	6	ABU27936 Protein e
15	164	12.5	205	6	ABU15639 Protein e
16	137	10.4	181	6	ABU18884 Protein e
17	133	10.1	202	6	ABU31958 Protein e
18	131	9.9	238	3	AAy70730 Klebsiell
19	130	9.9	227	6	ABU02540 S. pneumo
20	126.5	9.6	220	6	ABU21860 Protein e
21	126.5	9.6	230	5	ABp65425 Bifidobac
22	126	9.6	230	6	ABU49674 Protein e
23	125.5	9.5	241	7	ADc94690 E. faeciu
24	124.5	9.5	234	6	ABU46266 Protein e
25	124.5	9.5	232	6	ABU29712 Protein e

## ALIGNMENTS

### RESULT 1

AAy85635  
ID AAy85635 standard; protein; 241 AA.

XX AAy85635;

XX 07-FEB-2001 (first entry)

XX Antigen recognised by Ab capable of inducing G-CSF activity.

XX Antigenic protein; antibody; granulocyte colony stimulating factor;

XX G-CSF; cancer therapy; bone marrow suppression; mouse.

XX Mus sp.

XX WO200060075-A1.

XX 12-OCT-2000.

XX 31-MAR-2000; 2000WO-JP002080.

XX 01-APR-1999; 99JP-00095092.

XX (NIBS) JAPAN TOBACCO INC.

XX Sha S, Aoki Y, Nishi Y;

XX WPI; 2001-024452/03.

XX N-PSDB; AAC61149.

XX Gene encoding an antigen recognizing an antibody which induces granulocyte colony stimulating factor (G-CSF) expression for gene therapy and treatment of G-CSF associated disorders e.g. the side effects of cancer therapy.

XX Claim 2; Page 49-50; 58pp; Japanese.

XX The present invention relates to a gene encoding an antigenic protein recognised by an antibody or its fragments which can induce the production of granulocyte colony stimulating factor (G-CSF). Also included in the invention are partial sequences of the gene, antibodies recognising all or part of the antigenic protein, expression vectors containing the gene and host cells transformed by the vector. The gene is used for gene therapy, and compounds identified by screening using the gene sequence are used for treatment and prevention of disorders associated with G-CSF expression such as the side effects of cancer therapy (including bone marrow suppression). The present sequence represents the murine antigen of the invention

26	124.5	9.5	233	7	ADC95935	E. faeciu
27	123.5	9.4	236	6	ABU29217	Protein e
28	123	9.3	228	6	ABU47361	Protein e
29	123	9.3	234	5	ABP27712	Streptoco
30	123	9.3	234	6	ABU46430	Protein e
31	122.5	9.3	228	6	ABU25094	Protein e
32	122.5	9.3	231	6	ABU49976	Protein e
33	121	9.2	228	6	ABU48206	Protein e
34	120	9.1	228	6	ABU28874	Protein e
35	118	9.0	242	2	AAW22376	S. pneumo
36	111.5	8.5	212	6	ABU02694	S. pneumo
37	111	8.4	238	5	ABP27711	Streptoco
38	110.5	8.4	233	6	ABU23869	Protein e
39	109.5	8.3	243	6	ABM65009	Propionib
40	108.5	8.2	75	2	AAW21952	B6-bindin
41	108.5	8.2	76	2	AAW77663	HPV B6-bi
42	103	7.8	229	2	AAW53951	Bacillus
43	102.5	7.8	210	6	ABU30461	Protein e
44	101	7.7	213	4	AAU49973	Propionib
45	101	7.7	213	6	ABM46492	Propionib



XX SQ Sequence 241 AA;  
 Query Match 100.0%; Score 1317; DB 4; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-135;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTTGGGSLKHGNEIYIAPS 60  
 DB 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTTGGGSLKHGNEIYIAPS 60  
 QY 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
 DB 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
 QY 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKOLKERMAHAM 180  
 DB 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKOLKERMAHAM 180  
 QY 181 NEYPDSCAVLVRHGVVWGETWEKATMCCEYDYLFDIAVSMKMGLDPTQLPVGENGI 240  
 DB 181 NEYPDSCAVLVRHGVVWGETWEKATMCCEYDYLFDIAVSMKMGLDPTQLPVGENGI 240  
 QY 241 V 241  
 DB 241 V 241

RESULT 2  
 AAU77177  
 ID AAU77177 standard; protein; 241 AA.  
 AC AAU77177;  
 DT 02-JUL-2002 (first entry)  
 DE Murine G-CSF-inducible antibody binding protein, MM19.  
 KW Mouse; granulocyte-colony stimulating factor; G-CSF; MM19;  
 KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;  
 KW infection.  
 OS Mus sp.  
 PN WO200226978-A1.  
 PD 04-APR-2002.  
 PF 27-SEP-2001; 2001WO-JP008446.  
 PR 27-SEP-2000; 2000JP-00294191.  
 PA (NTSB) JAPAN TOBACCO INC.  
 PI Sha S, Mukai H, Aoki Y, Nishi Y;  
 DR WPI; 2002-340016/37.  
 DR N-PSDB; ABK47966.

PT Gene encoding protein binding to antibody having granulocyte-colony  
 PT stimulating factor (G-CSF) inducing activity, useful for screening  
 PT potential drugs treating G-CSF associated diseases.  
 PS Claim 1; Page 93-94; 103pp; Japanese.  
 CC The invention relates to a mouse or human gene (MM19) encoding a protein  
 CC which binds to antibodies or their fragments which induce granulocyte-  
 CC colony stimulating factor (G-CSF) secretion. The genes and proteins of  
 CC the invention are used in diagnosis, treatment and prevention of diseases  
 CC associated with G-CSF, including infections and neutrophil deficiency  
 CC disease. This sequence represents a mouse G-CSF-inducible antibody  
 CC binding protein, MM19

SQ Sequence 241 AA;  
 Query Match 100.0%; Score 1317; DB 5; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-135;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTTGGGSLKHGNEIYIAPS 60  
 DB 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTTGGGSLKHGNEIYIAPS 60  
 QY 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
 DB 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
 QY 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKOLKERMAHAM 180  
 DB 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKOLKERMAHAM 180  
 QY 181 NEYPDSCAVLVRHGVVWGETWEKATMCCEYDYLFDIAVSMKMGLDPTQLPVGENGI 240  
 DB 181 NEYPDSCAVLVRHGVVWGETWEKATMCCEYDYLFDIAVSMKMGLDPTQLPVGENGI 240  
 QY 241 V 241  
 DB 241 V 241

RESULT 3  
 AAU78361  
 ID AAU78361 standard; protein; 241 AA.  
 AC AAU78361;  
 DT 18-JUN-2002 (first entry)  
 DE Cell differentiation stimulator associated protein #2.  
 KW Cartilage cell differentiation stimulator; osteopathic;  
 KW Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;  
 KW membrane bound type transferrin-like protein; Mtf; cartilage disorder;  
 KW bone metabolism disease; cell differentiation; cell growth;  
 KW extracellular matrix related disease; mouse.  
 OS Mus sp.  
 PN JP2002020311-A.  
 PD 23-JAN-2002.  
 PF 07-JUL-2000; 2000JP-00206566.  
 PR 07-JUL-2000; 2000JP-00206566.  
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 PI WPI; 2002-287405/33.  
 DR N-PSDB; ABK12567.

PT A cartilage cell differentiation stimulator useful in the diagnosis of  
 PT biophylaxis, cell differentiation, cell growth and construction of  
 PT extracellular matrix related diseases.  
 PS Claim 2; Page 9-10; 17pp; Japanese.  
 CC The invention describes a cartilage cell differentiation stimulator  
 CC (containing a membrane-bound transferrin-like protein (Mtf-BP) and a  
 CC membrane bound type transferrin-like protein (Mtf)) and an animal-derived  
 CC concanavalin-like drug. The cartilage differentiation stimulator can be  
 CC used in diagnosis, prevention and treatment of cartilage and bone  
 CC metabolism diseases. They can also be used for diagnosing biophylaxis,  
 CC cell differentiation, cell growth and construction of extracellular  
 CC matrix related diseases. Mtf-BP strongly stimulates differentiation of  
 CC cartilage cells and exhibits similar action mechanism with that of plant

CC derived ConA. This is the amino acid sequence of a cartilage cell  
CC differentiation stimulator associated polypeptide described in the  
CC invention  
XX  
SQ Sequence 241 AA;  
Query Match 100.0%; Score 1317; DB 5; Length 241;  
Best Local Similarity 100.0%; Pred. No. 4.2e-135;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSGCOAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGCGISLKHGNEIYIAPS 60  
Db 1 MSGCOAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGCGISLKHGNEIYIAPS 60  
QY 61 GVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFMNAYTWRGAGAVIHTHSKAA 120  
Db 61 GVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFMNAYTWRGAGAVIHTHSKAA 120  
QY 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGYYRYDDMLVVPPIENTPEEKDLKERMAHAM 180  
Db 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGYYRYDDMLVVPPIENTPEEKDLKERMAHAM 180  
QY 181 NEYPDSCAVLVRHGVYVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240  
Db 181 NEYPDSCAVLVRHGVYVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240  
QY 241 V 241  
Db 241 V 241

RESULT 4  
AAG67127  
ID AAG67127 standard; protein; 242 AA.  
XX  
AC AAG67127;  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Amino acid sequence of a human enzyme.  
XX  
KW Human; enzyme; cancer; neurological disorder; epilepsy; stroke;  
KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;  
KW multiple sclerosis; Parkinson's disease; amyotrophic lateral sclerosis;  
KW meningitis; schizophrenia disorder; neuroskeletal disorder; allergy;  
KW Addison's disease; autoimmune disease; anemia; asthma; Crohn's disease;  
KW adult respiratory distress syndrome; atopic dermatitis; psoriasis;  
KW diabetes mellitus; osteoporosis; pancreatitis; rheumatoid arthritis;  
KW infection; genetic disorder; muscular dystrophy; Gaucher's disease;  
KW Huntington's chorea; sickle cell anemia; thalassemia; atherosclerosis;  
KW Von Willebrand's disease; Wilms' tumor; cell proliferative disorder;  
KW leukemia; hepatitis; cirrhosis; arteriosclerosis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FH Location/Qualifiers  
FT Modified-site 2 /note= "potential phosphorylation site"  
FT Modified-site 13 /note= "potential phosphorylation site"  
FT Modified-site 49 /note= "potential phosphorylation site"  
FT Modified-site 57 /note= "potential phosphorylation site"  
FT Modified-site 89 /note= "potential phosphorylation site"  
FT Modified-site 106 /note= "potential phosphorylation site"  
FT Modified-site 136 /note= "potential phosphorylation site"  
FT Modified-site 148 /note= "potential phosphorylation site"  
FT Modified-site 167 /note= "potential phosphorylation site"

FT Modified-site /note= "potential phosphorylation site"  
FT 209 /note= "potential phosphorylation site"  
FT 216 /note= "potential phosphorylation site"  
FT 223 /note= "potential phosphorylation site"  
FT /note= "potential phosphorylation site"  
WO200164896-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 01-MAR-2001; 2001WO-US006806.  
XX  
XX 01-MAR-2000; 2000US-0186307P.  
PR 28-MAR-2000; 2000US-0192532P.  
PR 30-MAR-2000; 2000US-0193578P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Tang YT, Lu DAM, Bandman O, Yue H, Azimzai Y, Lal P, Burford N;  
PI Baughn MR;  
XX WPI; 2001-550184/61.  
DR N-PSDB; AAH75155.  
XX  
PT Novel human enzyme molecule useful for treating and preventing, e.g.,  
PT cancer, genetic disorders, neurological disorders, autoimmune and  
PT inflammatory disorders.  
XX  
PS Claim 1; Page 117; 154pp; English.  
XX  
CC The present sequence represents a human enzyme. The enzyme polynucleotide  
CC and polypeptide are useful for diagnosis, treatment and prevention of  
CC cancers, neurological disorders (e.g. epilepsy, stroke, Alzheimer's  
CC disease, Pick's disease, Huntington's disease, dementia, multiple  
CC sclerosis, Parkinson's disease, amyotrophic lateral sclerosis, bacterial  
CC and viral meningitis, schizophrenic disorders (e.g. allergies, Addison's  
CC disorders), autoimmune/inflammatory disorders (e.g. atopic dermatitis, diabetes mellitus,  
CC anemia, asthma, Crohn's disease, atopic dermatitis, diabetes mellitus,  
CC osteoporosis, pancreatitis, psoriasis, rheumatoid arthritis, and viral,  
CC bacterial, fungal, parasitic, protozoal and helminthic infections),  
CC genetic disorder (e.g. Duchenne and Becker muscular dystrophy, Gaucher's  
CC disease, Huntington's chorea, sickle cell anemia, thalassemia, von  
CC Willebrand's disease and Wilms' tumor), and cell proliferative disorder  
CC (e.g. atherosclerosis, leukemia, hepatitis, cirrhosis, and  
CC arteriosclerosis). The polynucleotide is also useful in somatic or  
CC germline gene therapy  
SQ Sequence 242 AA;  
Query Match 94.6%; Score 1246.5; DB 4; Length 242;  
Best Local Similarity 94.2%; Pred. No. 2.2e-127;  
Matches 228; Conservative 9; Mismatches 4; Indels 1; Gaps 1;  
QY 1 MSGCOA-QGDCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGCGISLKHGNEIYIAP 59  
Db 1 MSGCDAREGDCSRRCGAQDKHEPRVLIPELCKQFYHLGWVTGCGISLKHGDEIYIAP 60  
QY 60 SGVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFMNAYTWRGAGAVIHTHSKA 119  
Db 61 SGVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFMNAYTWRGAGAVIHTHSKA 120  
QY 120 AVMATLLFPQGEFKITHQEMIKGIRKCTSGYYRYDDMLVVPPIENTPEEKDLKERMAHA 179  
Db 121 AVMATLLFPQGEFKITHQEMIKGIRKCTSGYYRYDDMLVVPPIENTPEEKDLKERMAHA 180  
QY 180 MNEYPDSCAVLVRHGVYVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 239  
Db 181 MNEYPDSCAVLVRHGVYVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240  
QY 240 IV 241



CC allowing gene inheritance to be studied through linkage analysis

XX  
SQ Sequence 242 AA;  
Query Match 93.8%; Score 1235.5; DB 2; Length 242;  
Best Local Similarity 93.8%; Pred. No. 3.4e-126;  
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MSGCOA-QGDCCSRPCGAQDKHPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYIAP 59  
DB 1 MSGCDANECDCCSRRCGAQDKHPRFLIPELCKQFVHLGWVTGTGGISLKHGDEIYIAP 60  
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSQCTPLFNAYTMRGAGAVIHTHSA 119  
DB 61 SGVQKERIQPEDMFVYDINEKDISGSPSKLKKSQCTPLFNAYTMRGAGAVIHTHSA 120  
QY 120 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHA 179  
DB 121 AVMATLLFPGRBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHA 180  
QY 180 MNEYDPSCAVLVRHGVYVWGVTWEKAKTMCCEYDYLFDIAVSMKMGGLDPTOLPVGNG 239  
DB 181 MNEYDPSCAVLVRHGVYVWGVTWEKAKTMCCEYDYLFDIAVSMKMGGLDPTOLPVGNG 240  
QY 240 IV 241  
DB 241 IV 242

## RESULT 7

AAV85636  
ID AAY85636 standard; protein; 242 AA.

XX AAY85636;

AC AAY85636;

XX 07-FEB-2001 (first entry)

DT Antigen recognised by Ab capable of inducing G-CSF activity.

DE Antigenic protein; antibody; granulocyte colony stimulating factor;

XX G-CSF; cancer therapy; bone marrow suppression; human.

XX Homo sapiens.

XX WO200060075-A1.

XX 12-OCT-2000.

XX 31-MAR-2000; 2000WO-JP002080.

XX 01-APR-1999; 99JP-00095092.

XX (NIBS) JAPAN TOBACCO INC.

XX Sha S, Aoki Y, Nishi Y;

XX WPI; 2001-024452/03.

XX N-PSDB; AAC61150.

XX Gene encoding an antigen recognizing an antibody which induces

XX granulocyte colony stimulating factor (G-CSF) expression for gene therapy

XX and treatment of G-CSF associated disorders e.g. the side effects of

XX cancer therapy.

XX Claim 3; Page 52-53; 58pp; Japanese.

XX The present invention relates to a gene encoding an antigenic protein

XX recognised by an antibody or its fragments which can induce the

XX production of granulocyte colony stimulating factor (G-CSF). Also

XX included in the invention are partial sequences of the gene, antibodies

XX recognising all or part of the antigenic protein, expression vectors

XX containing the gene and host cells transformed by the vector. The gene is

XX used for gene therapy, and compounds identified by screening using the

CC gene sequence are used for treatment and prevention of disorders  
CC associated with G-CSF expression such as the side effects of cancer  
CC therapy (including bone marrow suppression). The present sequence  
CC represents the human antigen of the invention

XX  
SQ Sequence 242 AA;

Query Match 93.8%; Score 1235.5; DB 4; Length 242;  
Best Local Similarity 93.8%; Pred. No. 3.4e-126;  
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MSGCOA-QGDCCSRPCGAQDKHPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYIAP 59  
DB 1 MSGCDANECDCCSRRCGAQDKHPRFLIPELCKQFVHLGWVTGTGGISLKHGDEIYIAP 60  
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSQCTPLFNAYTMRGAGAVIHTHSA 119  
DB 61 SGVQKERIQPEDMFVYDINEKDISGSPSKLKKSQCTPLFNAYTMRGAGAVIHTHSA 120  
QY 120 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHA 179  
DB 121 AVMATLLFPGRBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHA 180  
QY 180 MNEYDPSCAVLVRHGVYVWGVTWEKAKTMCCEYDYLFDIAVSMKMGGLDPTOLPVGNG 239  
DB 181 MNEYDPSCAVLVRHGVYVWGVTWEKAKTMCCEYDYLFDIAVSMKMGGLDPTOLPVGNG 240  
QY 240 IV 241  
DB 241 IV 242

## RESULT 8

AAU77178

ID AAU77178 standard; protein; 242 AA.

XX AAU77178;

AC AAU77178;

XX 02-JUL-2002 (first entry)

DT Human G-CSF-inducible antibody binding protein, MM19.

DE Human; granulocyte-colony stimulating factor; G-CSF; MM19;

XX antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;

XX infection.

XX Homo sapiens.

XX WO200226978-A1.

XX 04-APR-2002.

XX 27-SEP-2001; 2001WO-JP008446.

XX 27-SEP-2000; 2000JP-00294191.

XX (NIBS) JAPAN TOBACCO INC.

XX Sha S, Mukai H, Aoki Y, Nishi Y;

XX WPI; 2002-340016/37.

XX N-PSDB; ABK47967.

XX Gene encoding protein binding to antibody having granulocyte-colony

XX stimulating factor (G-CSF) inducing activity, useful for screening

XX potential drugs treating G-CSF associated diseases.

XX Claim 3; Page 96-97; 103pp; Japanese.

XX The invention relates to a mouse or human gene (MM19) encoding a protein

XX which binds to antibodies or their fragments which induce granulocyte-

XX colony stimulating factor (G-CSF) secretion. The gene and protein of

XX the invention are used in diagnosis, treatment and prevention of diseases

CC associated with G-CSF, including infections and neutrophil deficiency  
CC disease. This sequence represents a human G-CSF-inducible antibody  
CC binding protein, MMRL9  
XX  
SQ Sequence 242 AA;  
Query Match 93.8%; Score 1235.5; DB 5; Length 242;  
Best Local Similarity 93.8%; Pred. No. 3.4e-126;  
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;  
Yy 1 MSGCQA-QGDCSPCAQDKEHPRFLPELCKQFYHLGWVTGGGSLKHGNEIYIAP 59  
Db 1 MSGCDAWEGDCSRCAQDKEHPRFLPELCKQFYHLGWVTGGGSLKHGDEIYIAP 60  
Yy 60 SGVOKERIQEDMFVCDINEODISGPASKKLKKSOCTPLFMNAYTMRGAGAVIHTSKA 119  
Db 61 SGVOKERIQEDMFVCDINEODISGPSKSLKKSOCTPLFMNAYTMRGAGAVIHTSKA 120  
Yy 120 AVMATLFPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
Db 121 AVMATLFPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
Yy 180 MNEYPDSCAVLRHGVVWGETWEKAKTMCCEYDYLFDIAVSMKKGLDPTOLPVGNG 239  
Db 181 MNEYPDSCAVLRHGVVWGETWEKAKTMCCEYDYLFDIAVSMKKGLDPTOLPVGNG 240  
Yy 240 IV 241  
Db 241 IV 242  
RESULT 9  
ABB65485  
ID ABB65485 standard; protein; 227 AA.  
XX  
AC ABB65485;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 23247.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL09588.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 23247; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 227 AA;  
Query Match 59.7%; Score 786; DB 4; Length 227;  
Best Local Similarity 70.5%; Pred. No. 4.3e-77;  
Matches 148; Conservative 20; Mismatches 42; Indels 0; Gaps 0;  
Yy 21 EHRFLPELCKQFYHLGWVTGGGSLKHGNEIYIAPSGVOKERIQEDMFVCDINEQ 80  
Db 22 EHRFLPELCKQFYHLGWVTGGGSLKHGNEIYIAPSGVOKERIQEDMFVCDITGK 71  
Yy 81 DISPPASKKLKKSOCTPLFMNAYTMRGAGAVIHTSKAAVMTLLFPFGQEFKITHQEMI 140  
Db 72 DLQLPPEIKGLKKSOCTPLFMNAYTMRGAGAVIHTSKAAVMTLLFPFGQEFKITHQEMI 131  
Yy 141 KGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHANEYDSCAVLRHGVVWNG 200  
Db 132 KGVYDEADKRYLYRDEELVPIIENTPEERDLADSMYAAWMEYPCSAILVRRHGVVWNG 191  
Yy 201 ETWEKAKTMCCEYDYLFDIAVSMKKGLDLP 230  
Db 192 QNWEKAKTMCCEYDYLFDIAVSMKKGLDLP 221  
RESULT 10  
AAO10783  
ID AAO10783 standard; protein; 129 AA.  
XX  
AC AAO10783;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 24675.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US004927.  
XX  
PR 28-FEB-2000; 2000US-00515126.  
XX  
PR 18-MAY-2000; 2000US-00577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-514838/56.  
XX  
DR N-PSDB; AAI90714.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.  
XX  
PS Claim 20; SEQ ID NO 24675; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX  
SQ Sequence 129 AA;  
  
Query Match 18.8%; Score 248; DB 4; Length 129;  
Best Local Similarity 46.4%; Pred. No. 1.2e-18;  
Matches 51; Conservative 4; Mismatches 7; Indels 48; Gaps 2;  
  
QY 1 MSGCQA-QGDCSRPGQAQ-----18  
DB 20 MSGCNARKGDCSRRGSHLXNXPIDTDLNFFLTSAKVKCFPKKESIQTVVXSPG 79  
  
QY 19 -----DKHPRFLPELCKQFVHLGWVTGTGGISLKHGNEIYIAPSGV 62  
DB 80 KXMTYDKHPRVLIPLXCIQFVHLGWVTGTGGIILKHGDEIYIAPSGV 129  
  
RESULT 11  
AAG74374  
ID AAG74374 standard; protein; 59 AA.  
XX  
AC AAG74374;  
XX  
DT 03-SEP-2001 (first entry)  
DE Human colon cancer antigen protein SEQ ID NO:5138.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma.  
XX Homo sapiens.  
XX  
XX W0200122920-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 28-SEP-2000; 2000WO-US026524.  
XX  
XX 29-SEP-1999; 99US-0157137P.  
XX 03-NOV-1999; 99US-0163280P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
XX WPI; 2001-235357/24.  
XX DR N-PSDB; AAH33805.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers.  
XX  
XX Claim 11; Page 6841; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77798 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the  
XX proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene therapy  
XX and vaccine production. N and P may be used in the prevention, diagnosis  
XX and treatment of diseases associated with inappropriate P expression. For  
XX example, N and P may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of P by expressing inactive proteins or to  
XX supplement the patients own production of P. Additionally, N may be used  
XX to produce the colon cancer-associated P, by inserting the nucleic acids  
XX into a host cell and culturing the cell to express the proteins. N and P  
XX can be used in the prevention, diagnosis and treatment of colorectal  
XX carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent

CC sequences used in the exemplification of the present invention. N.B.  
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
XX to 1052, 7921 and 7922  
XX  
SQ Sequence 59 AA;  
  
Query Match 16.9%; Score 222; DB 4; Length 59;  
Best Local Similarity 95.2%; Pred. No. 2.8e-16;  
Matches 40; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 203 GETWEKAKTMCEYDYLFDIAVSMKMGDLPTQLPVGENGIV 241  
DB 18 GETWEKAKTMCEYDYLFDIAVSMKMGDLPTQLPVGENGIV 59  
  
RESULT 12  
ABU17451  
ID ABU17451 standard; protein; 212 AA.  
XX  
XX AC ABU17451;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by Prokaryotic essential gene #2978.  
DE  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Bacillus anthracis.  
XX  
XX W0200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA21321.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 45375; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 212 AA;  
Query Match 15.3%; Score 201; DB 6; Length 212;  
Best Local Similarity 28.8%; Pred. No. 3.5e-13;  
Matches 60; Conservative 29; Mismatches 93; Indels 26; Gaps 6;  
QY 29 ELCKQFYHL-----GWVTGGGSLKHGNE---IYAPSGVQKERIQPEDMFVC 75  
Db 3 QLFQWYDLSEIKKELTRNWPATSGNISIKVSHPELTLITAGKDKTKTTPDDFLV 62  
QY 76 DINEQDISGPPASKLLKKSQCTPLFMNAYTMRGAGAVIHTSKAAVMATLLPFGQEFKIT 135  
Db 63 D-----HLGVPLETELRPSAETILHTHIYNNWAGCVLHVHTDNNVITNLY-SDAVTLQ 117  
QY 136 HOEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYDPSCAVLVRHG 195  
Db 118 NQEIILKALDIWEGA-----TTHIPIENHAHPTLGENFRKHQ--GDSGAVLIRNHG 169  
QY 196 VYVWGTEWKAKTMCYCYDLFLDIAMV 223  
Db 170 ITWVGSRDSDFAKKRLAYEPLFQFIKL 197  
RESULT 13  
ABU41912  
ID ABU41912 standard; protein; 204 AA.  
XX  
AC ABU41912;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by prokaryotic essential gene #27439.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX *Pseudomonas syringae*.  
XX  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
XX 06-SEP-2001; 2001US-00948993.  
XX  
XX 25-OCT-2001; 2001US-0342923P.  
XX  
XX 08-FEB-2002; 2002US-00072851.  
XX  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA45782.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 69836; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 204 AA;  
Query Match 14.8%; Score 195; DB 6; Length 204;  
Best Local Similarity 25.2%; Pred. No. 1.5e-12;  
Matches 53; Conservative 36; Mismatches 93; Indels 28; Gaps 6;  
QY 21 EHRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAPSGVQKERIQPEDMFVC 77  
Db 12 EAGRFL-----YGRGWSPATSSNYSVRLSEALLTVSGHKHKGQGLGDDVLATDLA 63  
QY 78 NEQDISGPPASKLLKKSQ---CTPLFMNAYTMRGAGAVIHTSKAAVMATLLPFGQEFKI 134  
Db 64 NSLEPKKPSAETLLHTLQYLCP-----QVGAVLHTHSVNTATVLSRLTASDLV 114  
QY 135 THOEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYDPSCAVLVRH 194  
Db 115 EDVELQKAF-----NGVLTHESQVWVPIFNDQDIARLANVQVFLDAHPCECAGYLIRH 169  
QY 195 GVTVWGTEWKAKTMCYCYDLFLDIAMV 224  
Db 170 GLYTWCGRMSDALRQIEAFEFLECECLKMR 199  
RESULT 14  
ABU27936  
ID ABU27936 standard; protein; 204 AA.  
XX  
XX ABU27936;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by prokaryotic essential gene #13463.  
XX  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Enterobacter cloacae.

XX PN WO200277183-A2. 147 TSGGYRYDDMLVPIIETPEEKDKERMAHAMNEYDSCAVLVRGGVYVWGETWEKA 206

XX PD 125 --GQTHLDTVAIPFDNDQIDALASRIAHYAERPLNYGLRGLHGLTCWRDVAEA 181

XX PF 03-OCT-2002.

XX PP 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX XX WPI: 2003-029926/02.

XX DR N-PSDB; ACA31806.

XX DR New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX XX Claim 25; SEQ ID NO 55860; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

XX CC proliferation of an organism. The antisense nucleic acids are useful for

XX CC identifying proteins or screening for homologous nucleic acids required

XX CC for cellular proliferation to isolate candidate molecules for rational

XX CC drug discovery programs, or for screening homologous nucleic acids

XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX CC the target prokaryotic essential genes. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX XX Sequence 204 AA;

XX SQ Query Match 13.1%; Score 172; DB 6; Length 204;

XX SQ Best Local Similarity 25.0%; Pred. No. 4.8e-10;

XX SQ Matches 50; Conservative 40; Mismatches 99; Indels 12; Gaps 4;

QY 29 ELCKQFYHLGWVTGGGSLKHGNEI-YIAPSGVQKERIOPEDMFVCDINEQDISGPPA 87

DB 12 DACRWIGAKWAPATGGNMSVRQDEHLWLSSGKDKSLTADFL----QVEIATNRA 66

QY 88 SKLKKSSQCTPLFMNAYTM-RGAGAVIHTHSAKAAVNATLLPFGQFKITHQMIKIGKRC 146

DB 67 PSGRPRSAETGLHTLIYLFPPANAVLHVHTVNTVLTRLNKAELNLSGPFEMQSLT-- 124

QY 147 TSGGYRYDDMLVPIIETPEEKDKERMAHAMNEYDSCAVLVRGGVYVWGETWEKA 206

DB 125 --GQTHLDTVAIPFDNDQIDALASRIAHYAERPLNYGLRGLHGLTCWRDVAEA 181

QY 207 KTMCECYDYLEDIAVSMKKM 226

DB 182 RRHLEGLFLECEMRLQL 201

RESULT 15

ABU15639

ID ABU15639 standard; protein; 205 AA.

XX AC ABU15639;

XX XX 19-JUN-2003 (first entry)

XX DT Protein encoded by Prokaryotic essential gene #1166.

XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX KW Pseudomonas aeruginosa.

XX OS WO200277183-A2.

XX PN 03-OCT-2002.

XX PD 21-MAR-2002; 2002WO-US009107.

XX PF 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX XX (ELIT-) ELITRA PHARM INC.

XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX XX WPI: 2003-029926/02.

XX XX N-PSDB; ACA19509.

XX XX New antisense nucleic acids, useful for identifying proteins or screening

XX XX for homologous nucleic acids required for cellular proliferation to

XX XX isolate candidate molecules for rational drug discovery programs.

XX XX Claim 25; SEQ ID NO 43563; 1766pp; English.

XX XX The invention relates to an isolated nucleic acid comprising any one of

XX XX the 6213 antisense sequences given in the specification where expression

XX XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX XX encoding a polypeptide whose expression is inhibited by the antisense

XX XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX XX polypeptide or its fragment whose expression is inhibited by the

XX XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX XX proliferation; (7) identifying a compound that influences the activity of

XX XX the gene product or that has an activity against a biological pathway

XX XX required for proliferation, or that inhibits cellular proliferation; (8)

XX XX identifying a gene required for cellular proliferation or the biological

XX XX pathway in which a proliferation-required gene or its gene product lies

XX XX or a gene on which the test compound that inhibits proliferation of an

XX XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX XX compound's activity; (11) a culture comprising strains in which the gene

XX XX product is overexpressed or underexpressed; (12) determining the extent

XX XX to which each of the strains is present in a culture or collection of

XX XX strains; or (13) identifying the target of a compound that inhibits the

XX XX proliferation of an organism. The antisense nucleic acids are useful for

XX XX identifying proteins or screening for homologous nucleic acids required

XX XX for cellular proliferation to isolate candidate molecules for rational

XX XX drug discovery programs, or for screening homologous nucleic acids

XX XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

XX XX the target prokaryotic essential genes. Note: The sequence data for this

XX XX patent did not form part of the printed specification, but was obtained

XX XX in electronic format directly from WIPO at

XX XX ftp.wipo.int/pub/published\_pct\_sequences

XX XX Sequence 204 AA;

XX XX Query Match 13.1%; Score 172; DB 6; Length 204;

XX XX Best Local Similarity 25.0%; Pred. No. 4.8e-10;

XX XX Matches 50; Conservative 40; Mismatches 99; Indels 12; Gaps 4;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:32:21 ; Search time 22 seconds  
(without alignments)  
565.539 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSCCQAGDCSCSPCGAQDK.....SMKMGILDPTQLPVGENGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pdp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pdp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pdp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pdp.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pdp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	164	12.5	238	4	US-09-252-991A-30160	Sequence 30160, A
2	142	10.8	205	4	US-09-489-039A-13421	Sequence 13421, A
3	139	10.6	231	4	US-09-172-952-32	Sequence 32, Appl
4	131	9.9	238	4	US-09-172-952-18	Sequence 18, Appl
5	128.5	9.8	231	4	US-09-172-952-33	Sequence 33, Appl
6	125.5	9.5	241	4	US-09-107-532A-4317	Sequence 4317, Ap
7	124.5	9.5	233	4	US-09-107-532A-5562	Sequence 5562, Ap
8	123.5	9.4	241	4	US-09-134-000C-6065	Sequence 6065, Ap
9	118	9.0	242	2	US-08-472-534-2	Sequence 21, Appl
10	113.5	8.6	240	3	US-08-826-842B-21	Sequence 10152, A
11	109.5	8.3	225	4	US-09-489-039A-10152	Sequence 20, Appl
12	109.5	8.3	231	3	US-08-826-842B-20	Sequence 11, Appl
13	108.5	8.2	75	2	US-08-840-683-11	Sequence 11, Appl
14	108.5	8.2	75	2	US-08-555-722-11	Sequence 11, Appl
15	108.5	8.2	75	2	US-09-384-301-11	Sequence 12402, A
16	106	8.0	244	4	US-09-489-039A-8943	Sequence 8943, Ap
17	105.5	8.0	285	4	US-09-489-039A-12402	Sequence 14, Appl
18	105	8.0	229	3	US-08-926-842B-14	Sequence 9000, Ap
19	101.5	7.7	216	4	US-09-489-039A-9000	Sequence 11424, A
20	89.5	6.8	225	4	US-09-489-039A-11424	Sequence 13768, A
21	89.5	6.8	225	4	US-09-489-039A-13768	Sequence 15, Appl
22	85	6.5	462	4	US-09-129-112-15	Sequence 2, Appl
23	82.5	6.3	1729	4	US-09-533-690-2	Sequence 10, Appl
24	81.5	6.2	260	3	US-08-081-929-10	Sequence 8023, Ap
25	80.5	6.1	399	4	US-09-489-039A-8023	Sequence 2, Appl
26	80	6.1	234	1	US-08-850-118-2	Sequence 2, Appl
27	80	6.1	234	2	US-09-008-253-2	

28	80	6.1	234	3	US-09-093-335-2	Sequence 2, Appl
29	80	6.1	536	4	US-08-426-509A-12	Sequence 12, Appl
30	80	6.1	536	4	US-08-232-545-12	Sequence 12, Appl
31	80	6.1	536	5	PCT-US95-05008-12	Sequence 12, Appl
32	79	6.0	381	3	US-08-857-076-106	Sequence 106, Appl
33	79	6.0	558	4	US-09-252-991A-16908	Sequence 16908, A
34	79	6.0	1724	3	US-08-857-076-12	Sequence 12, Appl
35	78	5.9	544	4	US-09-614-912-72	Sequence 72, Appl
36	77.5	5.9	298	4	US-09-540-236-2717	Sequence 2717, Ap
37	76	5.8	537	4	US-08-426-509A-11	Sequence 11, Appl
38	76	5.8	537	4	US-08-232-545-11	Sequence 11, Appl
39	76	5.8	537	5	PCT-US95-05008-11	Sequence 11, Appl
40	76	5.8	567	3	US-08-841-483-2	Sequence 2, Appl
41	76	5.8	567	3	US-09-382-911-2	Sequence 2, Appl
42	75.5	5.7	307	2	US-08-216-894-6	Sequence 6, Appl
43	75.5	5.7	307	3	US-09-115-746-6	Sequence 6, Appl
44	75.5	5.7	502	3	US-08-123-934A-4	Sequence 4, Appl
45	75.5	5.7	502	4	US-09-874-628-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-30160  
; Sequence 30160, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30160  
; LENGTH: 238  
; TYPE: PPT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30160

Query Match	12.5%	Score 164;	DB 4;	Length 238;
Best Local Similarity	25.9%	Pred. No. 9.9e-10;		
Matches	51;	Conservative	28;	Mismatches 98; Indels 20; Gaps 5;
QY	24	RFLPELCKQFVHLGWVTGGISLKHGNE-IYIAPSGVQKRIQPEDMFVCDINEQDI	82	
DB	50	RFL-----YGRGWSPATSNYSARLDEQALLTVSGKHKGQGFDDVLA-----TDL	96	
QY	83	SGPPASKKLKKSOCTPLFMNAYTMGA-GAVITHSKAAVMATLLPPGQEFKITHQEMTK	141	
DB	97	AGNSLEPGKPAEFLITLQYVAMPALGAVLTHSVNATVLSRLVRGRLQDYELQK	156	
QY	142	GIRKCTSGGYRDMVLVVPPIENTPEEKDKERMAHMANEYDSCAVLRHGVYVWGE	201	
DB	157	AF-----AGVTHGEGVEVPIFDNDQDIARLASRVQVLEAHPCPGYLIRGHGLYTWGA	211	
QY	202	TWEAKTMCECDYLPD	218	
DB	212	RMSDALRQVEAEFLFE	228	

RESULT 2  
US-09-489-039A-13421  
; Sequence 13421, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLBBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 2709.2004001  
;; CURRENT APPLICATION NUMBER: US/09/489, 039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 13421  
;; LENGTH: 205  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13421

Query Match 10.8%; Score 142; DB 4; Length 205;  
Best Local Similarity 23.7%; Pred. No. 2e-07;  
Matches 45; Conservative 37; Mismatches 96; Indels 12; Gaps 4;  
QY 38 GWTGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLKKSQC 96  
Db 23 GWAPATGNNMSVRQDDTWCLWLSSEGRDKSLTTEDEL-----QVEIATNQAPSGRKPSE 77  
QY 97 TPLFMNAYTM-RGAGAVIHTHSAAVNATLLPFGQEFKITHQEMIKGIRKCTSGGYRYD 155  
Db 78 TGLHTLVYRLFPPEANVVLHVHTVNAVLSRIKSDTLALQGYEMQK-----TSLSGQSHL 132  
QY 156 DMLVVPPIENTPEKDLKERMAHMANEYDPSCAVLVRHGVVYVNGETWEKATMCECYD 215  
Db 133 DTVPALFDNDQDIDALAARIADYAOQTRPLRYGELLRGHLTGWKDIQARQLEGLEF 192  
QY 216 LFDIAVSMKK 225  
Db 193 LFECELMRRR 202

## RESULT 3

US-09-172-952-32  
;; Sequence 32, Application US/09172952  
;; Patent No. 6368793  
;; GENERAL INFORMATION:  
;; APPLICANT: Hoch, James  
;; APPLICANT: Dartois, Veronique  
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS  
;; FILE REFERENCE: 234/191  
;; CURRENT APPLICATION NUMBER: US/09/172,952  
;; CURRENT FILING DATE: 1998-10-14  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 32  
;; LENGTH: 231  
;; TYPE: PRT  
;; ORGANISM: Yias-Ec  
US-09-172-952-32

Query Match 10.6%; Score 139; DB 4; Length 231;  
Best Local Similarity 27.8%; Pred. No. 5.1e-07;  
Matches 65; Conservative 27; Mismatches 96; Indels 46; Gaps 13;  
QY 26 LIPELCKOYHLGWGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDIS 83  
Db 10 LAANLALPAHL--VTFWGNVSAVDETRWVVIKPSGVYDVMTADDMMVVVEI---AS 63  
QY 84 GPPASKLKKSQCTPLFMNAYTMRG-AGAVIHTHSAKAAVM---ATLLPFGQEFKITHQEM 139  
Db 64 GKVEGSKXPSSDTPTHALYRYAEIGGIVHTSHRATIWSQAGLDLPA--WGTHADY 121  
QY 140 IKGIRKCTS-----GGYRYDDMLVVPPIENTPEKDLKERMAHMANEYDPSCAVLVR 192  
Db 122 FYGALPCTQMTAEELINGEYEQ---TGEVLIETFEER-----GRSPAQIP---AVLVH 169  
QY 193 RHGVVYWG---ETWEKATMCEC--YDYLFDIAVSMKKGLDPTQLPVGENGIV 241  
Db 170 SHGPPAWGNKAADAVHNAVLECAWGLFSRQLA-----POLPAMQNEL 215

## RESULT 4

US-09-172-952-18  
;; Sequence 18, Application US/09172952  
;; Patent No. 6368793  
;; GENERAL INFORMATION:  
;; APPLICANT: Hoch, James  
;; APPLICANT: Dartois, Veronique  
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS  
;; FILE REFERENCE: 234/191  
;; CURRENT APPLICATION NUMBER: US/09/172,952  
;; CURRENT FILING DATE: 1998-10-14  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 18  
;; LENGTH: 238  
;; TYPE: PRT  
;; ORGANISM: Yias-Ko  
US-09-172-952-18

Query Match 9.9%; Score 131; DB 4; Length 238;  
Best Local Similarity 28.0%; Pred. No. 3.9e-06;  
Matches 60; Conservative 21; Mismatches 87; Indels 46; Gaps 11;  
QY 38 GWTGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLKKSQ 95  
Db 20 GLVTFWGNVSAVDETRWVVIKPSGVYDVMTADDMMVVEM-----ASGVVEGKKPSS 75  
QY 96 CTLEFNAY-TMRGAGAVIHTHSAKAAVM---ATLLPFGQEFKITHQEMIKGIRKCTS--- 148  
Db 76 DTPHIALYRYDQIGGIVHTSHRATIWSQAGLDLPA--WGTHADYFYGAIPCTRRMT 133  
QY 149 ---GGYRYDDMLVVPPIENTPEKDLKERMAHMANEYDPSCAVLVRHGVVYWG----- 200  
Db 134 VEEINGEYEQ---TGEVLIETFEERQGLDPA-----QIP---AVLVHSHGPPAWGNKAA 181  
QY 201 ETWEKATMCECYDYLFDIAVSMKKGLDPTQLP 234  
Db 182 DAVHNAVLEEC-----AYNGLFSRQMP 204

## RESULT 5

US-09-172-952-33  
;; Sequence 33, Application US/09172952  
;; Patent No. 6368793  
;; GENERAL INFORMATION:  
;; APPLICANT: Hoch, James  
;; APPLICANT: Dartois, Veronique  
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS  
;; FILE REFERENCE: 234/191  
;; CURRENT APPLICATION NUMBER: US/09/172,952  
;; CURRENT FILING DATE: 1998-10-14  
;; NUMBER OF SEQ ID NOS: 33  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 33  
;; LENGTH: 231  
;; TYPE: PRT  
;; ORGANISM: Yias-Hi  
US-09-172-952-33

Query Match 9.8%; Score 128.5; DB 4; Length 231;  
Best Local Similarity 27.8%; Pred. No. 7.1e-06;  
Matches 50; Conservative 23; Mismatches 74; Indels 33; Gaps 9;  
QY 35 YHLGWGTGGGISL--KHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLK 92  
Db 19 HHL--VTFWGNVSAIDREKLVVIKPSGVYDVMTENDMVVDL-----FTGNIVEGNKK 72  
QY 93 KSQCTPLFMNAY-TMRGAGAVIHTHSAKAAVMATLLPFGQEFKITHQEMIKGIRKCT--- 147  
Db 73 PSSDTPHLELYRQFPHIGGIVHTSHRATIWSQAGLDIIEVGTHTGDFYFTIPCTRQM 132  
QY 148 -----SGGYRYDDMLVVPPIENTPEKDLKERMAHMANEYDPSC-AVLVRHGVVYWG 201

Db 133 TTKKRGNY-----ELETKGVIVETFLSRGIE-----PDNIPAVLVHSHGPFANWK 178

RESULT 6

US-09-107-532A-4317

; Sequence 4317, Application US/09107532A

; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4317:

SEQUENCE CHARACTERISTICS:

LENGTH: 241 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...241

SEQUENCE DESCRIPTION: SEQ ID NO: 4317:

US-09-107-532A-4317

Query Match 9.5%; Score 125.5; DB 4; Length 241;

Best Local Similarity 26.1%; Pred. No. 1.6e-05;

Matches 59; Conservative 20; Mismatches 66; Indels 81; Gaps 11;

Qy 45 GGISLKHGN-----EVIAPSGVQKERIQPEDMFVCDIN-----RQDISGPPASKKL 91

Db 30 GLVKLWGNVSEVDRELGVIVIKPSGVREYECMQADQMVVTDLSGNIIEED-----SL 81

Qy 92 KKSQCTPLFMNAY--TWRGAGAVIHTHSKAAVWATLLFPQOE---PKITHQEMIKGIRKCT 147

Db 82 KPSSDLPTHVLVYQTFEDITAIHTHSTHSHVWMAQ--AGRDLPAYGTTDAFYGVKVFCT 139

Qy 148 SGGYRYDDMLVPIIENTPPE-----KDLKERMAHAMNEYDPDSCAVLVR 193

Db 140 R-----QLTKKEVREAYEVHTGNVIVETFKERKLDLP--NEVP---GVLVYG 180

Qy 194 HGIVYVWGTEWKA-----KTMCECYDILFD 218

Db 181 HGPFTWGDSPMKAVENSLILDEICLMAKENLINPNICEIPOVLLD 226

RESULT 7

US-09-107-532A-5562

; Sequence 5562, Application US/09107532A

; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5562:

SEQUENCE CHARACTERISTICS:

LENGTH: 233 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...233

SEQUENCE DESCRIPTION: SEQ ID NO: 5562:

US-09-107-532A-5562

Query Match 9.5%; Score 124.5; DB 4; Length 233;

Best Local Similarity 25.4%; Pred. No. 2e-05;

Matches 59; Conservative 27; Mismatches 101; Indels 45; Gaps 10;

Qy 26 LIPELCKQFYHL-----GWVTGGGIGISLKHGNEIY--IAPSGVQKERIQPEDMFVCD 76

Db 2 MLEQLKEVYQANLDLPKHGLVKYTWGNVSADFDPETRYFVIKPSGVSYEEMTADDMVVVD 61

Qy 77 INEQDISGPPASKKLKKSQCTPLFMNAY--TWRGAGAVIHTHSKAAVWATLLFPQOEFKI 134

Db 62 LDNHIIEG-----KLNPSSDPTFHAVLYRSPQIGGVHTHSTWATIWAQGLDVPAMGT 116

Qy 135 THQEMIKGIRKCT-----SGYRYDDMLVPIIENTPEEKDLKERMAHAMNEYPD 185

Db 117 THADTFYGVSPCARPLFQOEIDSG--YETFGKV--ILETFKRK-----ID 159

Qy 186 SCA---VLVRHGVVYVWGTEWKEKTCCECYDYILFDIAVSMKMGGLDPTQLP 234

Db 160 PLAPGVLLHGHGPFPTWCKDAQSAVMNAVVLDEVCKMNLFTQLNSFSEELP 211

RESULT 8  
US-09-134-000C-6065  
; Sequence 6065, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 6065  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6065  
Query Match 9.4%; Score 123.5; DB 4; Length 241;  
Best Local Similarity 31.8%; Pred. No. 2.6e-05;  
Matches 57; Conservative 17; Mismatches 82; Indels 23; Gaps 9;  
QY 37 LGWVTGGGIS--LKHGNEIYIAPSGVOKERIQPEDMFVCDINEODISGPPASKKLKKS 94  
DB 29 LGLVLTGWNVSINSLGIIIVKPSGVYQEMTKQMVVTDKQGLE-----TNALKS 84  
QY 95 QCTPLEFNWY--TWRGAGAVHTHSKAAV---MATLLFPQGEFKITHQEMIKGIRKCTSGG 150  
DB 85 SDFPLHYLYOKMPEIGATAHTSLNSVTWQAQGRALP--PYCTTHADAFYGAVPCTRA- 141  
QY 151 YRYDMLVPIIENTPPE--KOLKERMAMANEYDSC-AVLVRHGVVYVWGTEWKA 206  
DB 142 -----LSESEIKENVEETGKIVETF-HEQLDPLAVGVLVYGHGPTWGTWTEKA 193  
RESULT 9  
US-08-472-534-2  
; Sequence 2, Application US/08472534  
; Patent No. 5919620  
; GENERAL INFORMATION:  
; APPLICANT: Hamel, Josee  
; APPLICANT: Brodeur, Bernard R  
; APPLICANT: Martin, Denis  
; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,534  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr, James F  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: Slovace-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000

TELEFAX: 212-596-9090  
TELEX: 14-8367  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 242 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-472-534-2  
Query Match 9.0%; Score 118; DB 2; Length 242;  
Best Local Similarity 22.9%; Pred. No. 0.00011;  
Matches 44; Conservative 31; Mismatches 83; Indels 34; Gaps 8;  
QY 17 AQDKHPRFLPELCKQFYHLGWTGTGGGIGSLKHGNEIYIA-PSGVOKERIQPEDMFVC 75  
DB 2 SQDKLIREQICDVCHKWQLGWAAANDGNVRLDEDTILATPTGISKSFITPEKLVKL 61  
QY 76 DINEQDISGP-----PASKKLKKSQCTPLFMNAYTWR-GAGAVIHTHSKAAV-MATLLRPG 129  
DB 62 NLKGEILEABGYCPSSEIKMHIRC-----YEEEDVRSVVVHAHPPIATGFAHAHPL 114  
QY 130 QBFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKOLKERMAMANEY-PDSCA 188  
DB 115 DYSILIESAIVGAIPITPPG-----VPSWTEVPE-----AITPYLPDHDV 155  
QY 189 VLVRHGVVYVWG 200  
DB 156 MLENHGALTVG 167  
RESULT 10  
US-08-926-842B-21  
; Sequence 21, Application US/08926842B  
; Patent No. 6030807  
; GENERAL INFORMATION:  
; APPLICANT: Sa-No. 6030807ueira, Isabel  
; APPLICANT: de Lencastre, Herminia  
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/926,842B  
; FILING DATE: 10-SEP-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-089 N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO

```
; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
US-08-926-842B-21

Query Match      8.6%; Score 113.5; DB 3; Length 240;
Best Local Similarity 26.4%; Pred. No. 0.00032;
Matches 47; Conservative 20; Mismatches 84; Indels 27; Gaps 7;

QY 40 VTGTGGGSL--KHGNEIYIAPSGVOKERIQEDMFVCDINEQDISGPPASKKLKKSQCT 97
DB 22 VTLTGNVSAVDRERGVLVPSGVDYSMTADDMMVVVSL-----SGEVGHHKPPSDDT 77

QY 98 PLFNWY--TMRGAGAVIHTSK--AAVATLLFPQGEFKITHQEMIKGIRKCTS----- 148
DB 78 PTHRLLYQAFPTIGGIVHTSRHATINAAQOPATGCTTHADYFYGTIPCTRMTEAEI 137

QY 149 GGYRYDDMLVPIENTPEEKDLKERMAMHANEYDSCAVLVRHGVVVGSTWEKA 206
DB 138 NGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVHSHGPPAWGKNAEDA 183

RESULT 11
US-09-489-039A-10152
; Sequence 10152, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10152
; LENGTH: 225
; TYPE: PKT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10152

Query Match      8.3%; Score 109.5; DB 4; Length 225;
Best Local Similarity 22.9%; Pred. No. 0.00081;
Matches 50; Conservative 38; Mismatches 87; Indels 43; Gaps 10;

QY 27 IPELCKQFVHLGWVTGTGGISLKHGNEIYIAPSGVOKERIQEDMFVCDINEQDISGPP 86
DB 20 IITDCTEMTRGLNQTAGNSVRYQGGMLTPTGIPYEKLTEDKIVFIDADQGEHQGL 79

QY 87 ASKLLKKSQCTPLFNWYTM--GAGAVIHTSKAAVMATLL---PFGQEFKITHQEMIKG 142
DB 80 PSEWRFRHOA-----AYOTRDAQVAVNHAVHCTAVSILNRPPIAHYMI----- 125

QY 143 IRKCTSGGYRYDDMLVPIENTPEEKDLKERMAMHANEYDSCAVLVRHGVVVGST 202
DB 126 ---AAAG---NSIFCAPVA--TFGTRELSEHVAVALKHRK---ATLIQHGLIACEAS 173

QY 203 WKAKTMCSCYD-----YLFDTAVSMKKGLDPTQLPV 235
DB 174 LEKALWLAHEVEVLAQLYLSTLAIT-----DP--VPV 203

RESULT 12
US-08-926-842B-20
; Sequence 20, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:

; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
US-08-926-842B-21

Query Match      8.6%; Score 113.5; DB 3; Length 240;
Best Local Similarity 26.4%; Pred. No. 0.00032;
Matches 47; Conservative 20; Mismatches 84; Indels 27; Gaps 7;

QY 40 VTGTGGGSL--KHGNEIYIAPSGVOKERIQEDMFVCDINEQDISGPPASKKLKKSQCT 97
DB 22 VTLTGNVSAVDRERGVLVPSGVDYSMTADDMMVVVSL-----SGEVGHHKPPSDDT 77

QY 98 PLFNWY--TMRGAGAVIHTSK--AAVATLLFPQGEFKITHQEMIKGIRKCTS----- 148
DB 78 PTHRLLYQAFPTIGGIVHTSRHATINAAQOPATGCTTHADYFYGTIPCTRMTEAEI 137

QY 149 GGYRYDDMLVPIENTPEEKDLKERMAMHANEYDSCAVLVRHGVVVGSTWEKA 206
DB 138 NGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVHSHGPPAWGKNAEDA 183

RESULT 11
US-09-489-039A-10152
; Sequence 10152, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10152
; LENGTH: 225
; TYPE: PKT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10152

Query Match      8.3%; Score 109.5; DB 3; Length 231;
Best Local Similarity 26.1%; Pred. No. 0.00084;
Matches 47; Conservative 19; Mismatches 83; Indels 31; Gaps 8;

QY 40 VTGTGGGSL--KHGNEIYIAPSGVOKERIQEDMFVCDINEQDISGPPASKKLKKSQCT 97
DB 22 VTLTGNVSAVDRERGVLVPSGVDYSMTADDMMVVVSL---TGEVWEGAKKPPSDDT 77

QY 98 PLFNWY--TMRGAGAVIHTSKAAVMATLLPFGQEFK---ITHQEMIKGIRKCTS----- 148
DB 78 PTHRLLYQAFPSIGGIVHTSRHATINAAQ--AGQSIPATGTTTHADYFYGTIPCTRMTEA 135

QY 149 --GGYRYDDMLVPIENTPEEKDLKERMAMHANEYDSCAVLVRHGVVVGSTWEKA 206
DB 136 EINGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVHSHGPPAWGKNAEDA 183

RESULT 13
US-08-840-683-11
; Sequence 11, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-840-683-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

RESULT 14
US-08-555-722-11
; Sequence 11, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-555-722-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

RESULT 15
US-09-384-301-11
; Sequence 11, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/384,301
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-09-384-301-11

Query Match      8.2%; Score 108.5; DB 3; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

Search completed: June 1, 2004, 13:35:58
Job time : 23 secs
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-840-683-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

RESULT 14
US-08-555-722-11
; Sequence 11, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-555-722-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

RESULT 15
US-09-384-301-11
; Sequence 11, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/384,301
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-09-384-301-11

Query Match      8.2%; Score 108.5; DB 3; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGSLKHG-----NEIYIAPSGVQK 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGEIYIAPSGCKR 49

Search completed: June 1, 2004, 13:35:58
Job time : 23 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:33:26 ; Search time 48 Seconds  
(without alignments)

1402.065 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCSRPCGAQDK.....SMKKMGIDPTQLPVGNGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1317	100.0	241	12	US-10-381-710-2
2	1317	100.0	241	16	US-10-629-329A-4
3	1246.5	99.6	242	15	US-10-220-381-2
4	1239.5	94.1	242	16	US-10-629-329A-2
5	1235.5	93.8	242	12	US-10-381-710-4
6	550	41.8	256	12	US-10-424-599-157170
7	548.5	41.6	524	12	US-10-425-114-64486
8	540	41.0	497	12	US-10-425-114-65135
9	535	40.6	522	12	US-10-424-599-273717
10	535	40.6	540	12	US-10-425-114-46271
11	458.5	34.8	459	12	US-10-425-114-61505
12	405	30.8	533	12	US-10-425-114-57875
13	315	23.9	64	14	US-10-029-386-28983
14	222	16.9	59	14	US-10-106-698-5148
15	203	15.4	202	14	US-10-156-761-14187

16	201	15.3	212	12	US-10-282-122A-45375	Sequence 45375, A
17	195	14.8	204	12	US-10-282-122A-59836	Sequence 59836, A
18	172	13.1	204	12	US-10-282-122A-55860	Sequence 55860, A
19	164	12.5	205	12	US-10-282-122A-43563	Sequence 43563, A
20	161.5	12.3	241	14	US-10-156-761-13948	Sequence 13948, A
21	142	10.8	104	12	US-10-424-599-183312	Sequence 183312, A
22	139	10.6	231	10	US-09-557-796-32	Sequence 32, Appl
23	137	10.4	181	12	US-10-282-122A-46808	Sequence 46808, A
24	133	10.1	202	12	US-10-282-122A-59882	Sequence 59882, A
25	131	9.9	238	10	US-09-557-796-18	Sequence 33, Appl
26	128.5	9.8	231	10	US-09-557-796-33	Sequence 49784, A
27	126.5	9.6	220	12	US-10-282-122A-49784	Sequence 49784, A
28	126	9.6	230	12	US-10-282-122A-77598	Sequence 77598, A
29	125	9.5	234	12	US-10-282-122A-74190	Sequence 74190, A
30	124.5	9.5	232	12	US-10-282-122A-57636	Sequence 57636, A
31	123.5	9.4	236	12	US-10-282-122A-57141	Sequence 57141, A
32	123	9.3	228	12	US-10-282-122A-75285	Sequence 75285, A
33	123	9.3	234	12	US-10-282-122A-74354	Sequence 74354, A
34	122.5	9.3	228	12	US-10-282-122A-53018	Sequence 53018, A
35	122.5	9.3	231	12	US-10-282-122A-77900	Sequence 77900, A
36	121	9.2	228	12	US-10-282-122A-76130	Sequence 76130, A
37	120	9.1	228	12	US-10-282-122A-56798	Sequence 56798, A
38	110.5	8.4	233	12	US-10-282-122A-51793	Sequence 51793, A
39	102.5	7.6	210	12	US-10-282-122A-58385	Sequence 58385, A
40	99.5	7.6	212	12	US-10-282-122A-48161	Sequence 48161, A
41	99	7.5	213	12	US-10-282-122A-50759	Sequence 50759, A
42	98.5	7.5	217	12	US-10-282-122A-61813	Sequence 61813, A
43	98	7.4	228	12	US-10-282-122A-73117	Sequence 73117, A
44	98	7.4	307	12	US-10-335-977-9353	Sequence 9353, Ap
45	98	7.4	315	12	US-10-335-977-9354	Sequence 9354, Ap

ALIGNMENTS

RESULT 1  
US-10-381-710-2  
; Sequence 2, Application US/10381710  
; Publication No. US20040052789A1  
; GENERAL INFORMATION:  
; APPLICANT: SHA, Shiken et al.  
; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME  
; FILE REFERENCE: 0230-0198P  
; CURRENT APPLICATION NUMBER: US/10/381,710  
; CURRENT FILING DATE: 2003-09-16  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Mouse macrophage cell RAW 264.7  
US-10-381-710-2

Query Match 100.0%; Score 1317; DB 12; Length 241;  
Best Local Similarity 100.0%; Pred. No. 1e-133;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSGCQAQGCCSRPCGAQDKHEPRFLPELCKQFYHLGVTGTGGISLKHGNEYIYAPS	60
Db	1	MSGCQAQGCCSRPCGAQDKHEPRFLPELCKQFYHLGVTGTGGISLKHGNEYIYAPS	60
Qy	61	GVQKERIQEDPMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKAA	120
Db	61	GVQKERIQEDPMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKAA	120
Qy	121	VMATLLFPQGFKTHQEMIKIRKCTSGGYRVDMLVPIENTPREKDKERMAHAM	180
Db	121	VMATLLFPQGFKTHQEMIKIRKCTSGGYRVDMLVPIENTPREKDKERMAHAM	180
Qy	181	NEYPDSCAVLRHGVYVWGTEWAKTMCYDYLFDIAVSMKKMGIDPTQLPVGNGI	240
Db	181	NEYPDSCAVLRHGVYVWGTEWAKTMCYDYLFDIAVSMKKMGIDPTQLPVGNGI	240



QY 241 V 241  
Db 241 V 241

## RESULT 2

US-10-629-329A-4  
; Sequence 4, Application US/10629329A  
; Publication No. US20040086848A1  
; GENERAL INFORMATION:  
; APPLICANT: DARNAY, BRYANT G.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND  
; TITLE OF INVENTION: POLYPEPTIDES OF RANK-ASSOCIATED INHIBITOR (RAIN)  
; FILE REFERENCE: UTSC:761US  
; CURRENT APPLICATION NUMBER: US/10/629,329A  
; CURRENT FILING DATE: 2003-07-29  
; PRIOR APPLICATION NUMBER: 60/399,205  
; PRIOR FILING DATE: 2002-07-29  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-629-329A-4

Query Match 100.0%; Score 1317; DB 16; Length 241;  
Best Local Similarity 100.0%; Pred. No. 1e-133; Indels 0; Gaps 0;  
Matches 241; Conservative 0; Mismatches 0;  
QY 1 MSGCQAQDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYIAPS 60  
Db 1 MSGCQAQDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYIAPS 60  
QY 61 GVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
Db 61 GVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAA 120  
QY 121 VMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180  
Db 121 VMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180  
QY 181 NEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENGI 240  
Db 181 NEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENGI 240  
QY 241 V 241  
Db 241 V 241

## RESULT 3

US-10-220-381-2  
; Sequence 2, Application US/10220381  
; Publication No. US20030207430A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: LU, Dyung Aina M.  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: YUE, Henry  
; APPLICANT: AZIMZAI, Yalda  
; APPLICANT: LAL, Preeti  
; APPLICANT: BURFORD, Neil  
; APPLICANT: BAUGHN, Mariah R.  
; TITLE OF INVENTION: HUMAN ENZYME MOLECULES  
; FILE REFERENCE: PF-0763 PCT  
; CURRENT APPLICATION NUMBER: US/10/220,381  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PERL Program  
; SEQ ID NO 2  
; LENGTH: 242

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030207430A1 2116390CD1  
US-10-220-381-2

Query Match 94.1%; Score 1246.5; DB 15; Length 242;  
Best Local Similarity 94.2%; Pred. No. 4.4e-126;  
Matches 228; Conservative 9; Mismatches 4; Indels 1; Gaps 1;  
QY 1 MSGCQA-QGDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYIAP 59  
Db 1 MSGCDAREGDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGDEIYIAP 60  
QY 60 SGVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKA 119  
Db 61 SGVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKA 120  
QY 120 AVMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
Db 121 AVMATLLPFGREFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
QY 180 MNEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENG 239  
Db 181 MNEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENG 240  
QY 240 IV 241  
Db 241 IV 242

## RESULT 4

US-10-629-329A-2  
; Sequence 2, Application US/10629329A  
; Publication No. US20040086848A1  
; GENERAL INFORMATION:  
; APPLICANT: DARNAY, BRYANT G.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND  
; TITLE OF INVENTION: POLYPEPTIDES OF RANK-ASSOCIATED INHIBITOR (RAIN)  
; FILE REFERENCE: UTSC:761US  
; CURRENT APPLICATION NUMBER: US/10/629,329A  
; CURRENT FILING DATE: 2003-07-29  
; PRIOR APPLICATION NUMBER: 60/399,205  
; PRIOR FILING DATE: 2002-07-29  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-629-329A-2

Query Match 94.1%; Score 1239.5; DB 16; Length 242;  
Best Local Similarity 93.8%; Pred. No. 2.5e-125;  
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;  
QY 1 MSGCQA-QGDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYIAP 59  
Db 1 MSGCDAGEGDCSRRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGSLKHGDEIYIAP 60  
QY 60 SGVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKA 119  
Db 61 SGVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKA 120  
QY 120 AVMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
Db 121 AVMATLLPFGREFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
QY 180 MNEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENG 239  
Db 181 MNEYDPDSCAVLVRHGVYVWGTEWKAETMCCEYDYLFDIAVSMKMGDLPTQLPVGENG 240

QY 240 IV 241  
Db 241 IV 242

RESULT 5  
US-10-381-710-4  
; Sequence 4, Application US/10381710  
; Publication No. US20040052789A1  
; GENERAL INFORMATION:  
; APPLICANT: SHA, Shiken et al.  
; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME  
; FILE REFERENCE: 0230-0198P  
; CURRENT APPLICATION NUMBER: US/10/381,710  
; CURRENT FILING DATE: 2003-09-16  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-381-710-4

Query Match 93.8%; Score 1235.5; DB 12; Length 242;  
Best Local Similarity 93.8%; Pred. No. 6.8e-125;  
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCOA-QGCCSRPCGAQDKHPRFLIPELCKQFVHLGWVTGGGSIKKGNEIYIAP 59  
Db 1 MSGCDAWEGDCCSRRCGAQDKHPRFLIPELCKQFVHLGWVTGGGSIKKGDEIYIAP 60

QY 60 SGVQKRIQPEDMFVCDINEODISGPPASKKLSQCTPLFMNAYTMRGAGAVIHTSKA 119  
Db 61 SGVQKRIQPEDMFVCDINEODISGPPASKKLSQCTPLFMNAYTMRGAGAVIHTSKA 120

QY 120 AVNATLLFPQEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEKOLKERMAMA 179  
Db 121 AVNATLLFPQEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEKOLKERMAMA 180

QY 180 MNEYPSCAVLRRHGVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGDPTOLPYGNG 239  
Db 181 MNEYPSCAVLRRHGVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGDPTOLPYGNG 240

QY 240 IV 241  
Db 241 IV 242

RESULT 6  
US-10-424-599-157170  
; Sequence 157170, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 157170  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_112946C.1.pep  
US-10-424-599-157170

Query Match 41.8%; Score 550; DB 12; Length 256;  
Best Local Similarity 47.6%; Pred. No. 1e-50;

Matches 121; Conservative 33; Mismatches 74; Indels 26; Gaps 9;

QY 1 MSGCOA-QGCCSRPC---GAQDKHPRFLIPELCKQFVHLGWVTGGGSIK-KGNE-- 54  
Db 1 MAATATADSVSSQAYLEGNVKE--TRILISELCRHFYSGLWVSGTSGSITIKVHDDSI 59

QY 55 -----IYIAPSGVQKRIQPEDMFVCDINEODISGPPAS-----KKLKESQCTPLFMNAYT 105  
Db 60 KPHQILIMSPSGVQKERMEDMYVLSHTGVSLSAPSPKPYPHKPKKDCGCLFMKAYE 119

QY 106 MRGAGAVIHTSKAAVMAATLLFP--QGEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIE 164  
Db 120 MCDAGAVIHTSHGIESCLVTLNPLAKEFKITHQEMIKGIK---GHHY---DELVIPIIE 173

QY 165 NTPPEKOLKERMAMANEYPDSCAVLRRHGVYVWGETWEKAKTWCBCYDYLFDIAVSMK 224  
Db 174 NTAYEYELTESLAKAIEAYPKTTAVLVRNHGIYIWDSDWISAKTQAECHYHLYLDAIRLH 233

QY 225 KMGLD---PTQLPV 235  
Db 234 QLGLDWSTPNHVPI 247

RESULT 7  
US-10-425-114-64486  
; Sequence 64486, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 64486  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-104-F8\_FLI.pep  
US-10-425-114-64486

Query Match 41.6%; Score 548.5; DB 12; Length 524;  
Best Local Similarity 45.6%; Pred. No. 4.1e-50;  
Matches 115; Conservative 36; Mismatches 68; Indels 33; Gaps 6;

QY 2 SGCOA-QGCCSRPCGAQDKHPRFLIPELCKQFVHLGWVTGGGSIKKG 52  
Db 11 SGCS-----CEAAVGWMASEAYLEGAPVREARELVNELCRHFAQGWVTGGGSIYKVN 65

QY 53 NE-----IYIAPSGVQKRIQPEDMFVCDINEODISGPPA-----SKLKESQCTPLF 100  
Db 66 DPTVPLADRLIVMSPSGVQKERMVADMYMAADGKVLAPVAKPWPKNPKKCTDCAPLF 125

QY 101 MNATYMRGAGAVIHTSKAAVMAATLLFPG--QEPKITHQEMIKGIRKCTSGGYRYDDMLV 159  
Db 126 MKAYLMRGAGAVIHTSHGIESCLVTLNPLAKEFKITHQEMIKGIKG-----HOYHDELV 179

QY 160 VPIIENTPEKOLKERMAMANEYPDSCAVLRRHGVYVWGETWEKAKTWCBCYDYLFDI 219  
Db 180 IPIIENTPEYELTDSLSEALAAYPKATAVLVRNHGIYVWGESWINAKTQAECHYHLYLDA 239

QY 220 AVSMKMGDPT 231  
Db 240 CIKLYQLGIDWT 251

```
RESULT 8
US-10-425-114-65135
; Sequence 65135, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65135
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMELMO17103D03_FLI.pep
US-10-425-114-65135

Query Match 41.0%; Score 540; DB 12; Length 497;
Best Local Similarity 48.5%; Pred. No. 3.2e-49;
Matches 111; Conservative 37; Mismatches 61; Indels 20; Gaps 5;

QY 16 GAQDKHPRFLPELCKQFHLGWVTGGGSIHLKHGNE-----IYIAPSGVQKRIQEDMFVC 67
DB 3 GAVPRE-ARELVLCRHFYQAQWVTGGTGGTITVKVNDPAVPLADRLIYVSPSGVQKERM 61

QY 68 QPEDMFVCDINEQDISGPPA-----SKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMA 123
DB 62 VAEDMYVMAADGKVSAPVAKPWPKPCKTDCAPLFMAKAYLMRGAGAVIHTSHGIETCIA 121

QY 124 TLLFPG-QEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKDKERMAHAMNE 182
DB 122 TMLIPGAKEFRVTHMEMIKIGK-----HGYHDELVIPIIENTPYEYELTDSLSEAIAA 175

QY 183 YPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKMGDPT 231
DB 176 YPRATAVLVRHGVYVWGESWINAKTOACBCHYLLDACIKLYQLGIDWT 224

RESULT 9
US-10-424-599-273717
; Sequence 273717, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 273717
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_89188C.1.pep
US-10-424-599-273717

Query Match 40.6%; Score 535; DB 12; Length 522;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

US-10-425-114-46271
; Sequence 46271, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46271
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701207771_FLI.pep
US-10-425-114-46271

Query Match 40.6%; Score 535; DB 12; Length 540;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

QY 24 RFLIPELCKQFHLGWVTGGGSIHLKHGNE-----IYIAPSGVQKRIQEDMFVC 75
DB 47 RALMAELCRHFYTLGWVTGGTGGSIHMKVHDDSIIPRQOQLILMAPSGVQKERMEDMYVL 106

QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAIILFP-GQ 130
DB 107 SHSGSVLSAPSAPKPPWPKPKSCDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 166

QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 167 EFRITHMEMIKGK-CHGY-DELVVPPIENTAYEYQLTESFAKAIEDYPKATAVL 220

QY 191 VRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKMGD-PTQLPV 235
DB 221 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSVTPNHGPI 268

RESULT 11
US-10-425-114-61505
; Sequence 61505, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

```
QY 24 RFLIPELCKQFHLGWVTGGGSIHLKHGNE-----IYIAPSGVQKRIQEDMFVC 75
DB 29 RALMAELCRHFYTLGWVTGGTGGSIHMKVHDDSIIPRQOQLILMAPSGVQKERMEDMYVL 88
QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAIILFP-GQ 130
DB 89 SHSGSVLSAPSAPKPPWPKPKSCDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 148
QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 149 EFRITHMEMIKGK-CHGY-DELVVPPIENTAYEYQLTESFAKAIEDYPKATAVL 202
QY 191 VRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKMGD-PTQLPV 235
DB 203 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSVTPNHGPI 250
```

```
RESULT 10
US-10-425-114-46271
; Sequence 46271, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46271
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701207771_FLI.pep
US-10-425-114-46271

Query Match 40.6%; Score 535; DB 12; Length 540;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

QY 24 RFLIPELCKQFHLGWVTGGGSIHLKHGNE-----IYIAPSGVQKRIQEDMFVC 75
DB 47 RALMAELCRHFYTLGWVTGGTGGSIHMKVHDDSIIPRQOQLILMAPSGVQKERMEDMYVL 106

QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAIILFP-GQ 130
DB 107 SHSGSVLSAPSAPKPPWPKPKSCDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 166

QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 167 EFRITHMEMIKGK-CHGY-DELVVPPIENTAYEYQLTESFAKAIEDYPKATAVL 220

QY 191 VRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKMGD-PTQLPV 235
DB 221 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSVTPNHGPI 268
```

APPLICANT: Cao, Yongwei  
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53313)B  
 CURRENT APPLICATION NUMBER: US/10/425,114  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 73128  
 SEQ ID NO 61505  
 LENGTH: 459  
 TYPE: PRT  
 ORGANISM: Zea mays  
 FEATURE:  
 OTHER INFORMATION: Clone ID: LIB143-005-F3\_FLI.pbp  
 US-10-425-114-61505

Query Match 34.8%; Score 458.5; DB 12; Length 459;  
 Best Local Similarity 50.0%; Pred. No. 1.8e-40;  
 Matches 91; Conservative 31; Mismatches 49; Indels 11; Gaps 3;

QY 55 IVAPSGVQKRIQPEDMFVCDINEQDISGPPA----SKLKKSQCTPLFMAYTMRGAG 110  
 DB 11 IVMSPSGVQKRWAEADMTVMAADGVLSPAPVAKPWNPKCTDCAPLFMKAYLMRGAG 70  
 QY 111 AVITHSKAAVMATLLPG-OBPKITHQEMIKGIRKCTSGGYRYDDMLVPIIENPTPE 169  
 DB 71 AVITHSHGIETCIATMLIPGAKERFVTHMEMIKGIG-----HGVDHDELVIPIIENPTPE 124  
 QY 170 KOLKERMAMNEYPDSCAVLVRHGHVYVWGETWEKAKTWCRCYDVLFDIAVSMKMGGLD 229  
 DB 125 YELTDSLSEAIAYPKATAVLVRHGHVYVWGESWINAKTQACVHYLLDACIKLYQLGID 184  
 QY 230 PT 231  
 DB 185 WT 186

RESULT 12  
 US-10-425-114-57875  
 Sequence 57875, Application US/10425114  
 Publication No. US20040034888A1  
 GENERAL INFORMATION:  
 APPLICANT: Liu, Jingdong  
 APPLICANT: Zhou, Yihua  
 APPLICANT: Kovalic, David K.  
 APPLICANT: Screen, Steven E.  
 APPLICANT: Tabaska, Jack E.  
 APPLICANT: Cao, Yongwei  
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53313)B  
 CURRENT APPLICATION NUMBER: US/10/425,114  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 73128  
 SEQ ID NO 57875  
 LENGTH: 533  
 TYPE: PRT  
 ORGANISM: Zea mays  
 FEATURE:  
 OTHER INFORMATION: Clone ID: UC-ZMFLM017051F08\_FLI.pbp  
 US-10-425-114-57875

Query Match 30.8%; Score 405; DB 12; Length 533;  
 Best Local Similarity 40.0%; Pred. No. 1.4e-34;  
 Matches 88; Conservative 32; Mismatches 64; Indels 36; Gaps 6;

QY 42 GTGGG-----ISLKHGNEIYIAPSGVQKRI--QPEDMF-----VCD- 76  
 DB 47 GAGGGSALLPAGHGHRHRRHQGRPRRAARRPHRYVTLRCAGEDSGSRHVCVG 106  
 QY 77 ----INEQDISGPPASKLKKSQCTPLFMAYTMRGAGAVITHSKAAVMATLLPG-GE 131  
 DB 107 CRWEVLSAPVAKPWNPKCTDCAPLFMKAYLMRGAGAVITHSHGIETCIATMLIPGAK 166

QY 132 FKITHQEMIKGIRKCTSGGYRYDDMLVPIIENPTPEKDKERMAMNEYPDSCAVL 191  
 DB 167 FRVTHMEMIKGIG-----HGVDHDELVIPIIENPTPEYELTDSLSEAIAYPKATAVL 220  
 QY 192 RPHGVYVWGETWEKAKTWCRCYDVLFDIAVSMKMGGLDPT 231  
 DB 221 RPHGVYVWGESWINAKTQACVHYLLDACIKLYQLGIDWT 260

RESULT 13  
 US-10-029-386-28983  
 Sequence 28983, Application US/10029386  
 Publication No. US20030194704A1  
 GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
 TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 FILE REFERENCE: AEOMICA-X-2  
 CURRENT APPLICATION NUMBER: US/10/029,386  
 CURRENT FILING DATE: 2001-12-20  
 NUMBER OF SEQ ID NOS: 34288  
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 28983  
 LENGTH: 64  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO CHR11.1  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.86  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.86  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.67  
 OTHER INFORMATION: SWISSPROT HIT: P47095, EVALUE 1.00e-11  
 US-10-029-386-28983

Query Match 23.9%; Score 315; DB 14; Length 64;  
 Best Local Similarity 98.3%; Pred. No. 3.6e-26;  
 Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 152 YRYDDMLVPIIENPTPEKDKERMAMNEYPDSCAVLVRHGHVYVWGETWEKAKTM 209  
 DB 7 YRYDDMLVPIIENPTPEKDKERMAMNEYPDSCAVLVRHGHVYVWGETWEKAKTM 64

RESULT 14  
 US-10-106-698-5148  
 Sequence 5148, Application US/10106698  
 Publication No. US20030109690A1  
 GENERAL INFORMATION:  
 APPLICANT: Ruben et al.  
 TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
 FILE REFERENCE: P4005F1  
 CURRENT APPLICATION NUMBER: US/10/106,698  
 CURRENT FILING DATE: 2002-03-27  
 PRIOR APPLICATION NUMBER: PCT/US00/26524  
 PRIOR FILING DATE: 2000-09-28  
 PRIOR APPLICATION NUMBER: US 60/157,137  
 PRIOR FILING DATE: 1999-09-29  
 PRIOR APPLICATION NUMBER: US 60/163,280  
 PRIOR FILING DATE: 1999-11-03  
 NUMBER OF SEQ ID NOS: 8564  
 SOFTWARE: PatentIn Ver. 3.0  
 SEQ ID NO 5148  
 LENGTH: 59  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-106-698-5148

Query Match 16.9%; Score 222; DB 14; Length 59;  
 Best Local Similarity 95.2%; Pred. No. 3.6e-16;  
 Matches 40; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 200 GETWEKAKTMCCEYDYLFDIAVSMKMGDLDPDTQLPVGENGIV 241  
Db 18 GETWEKAKTMCCEYDYLFDIAVSMKMGDLDPDPSQLPVGENGIV 59

RESULT 15  
US-10-156-761-14187  
; Sequence 14187, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 14187  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-14187

Query Match 15.4%; Score 203; DB 14; Length 202;  
Best Local Similarity 27.7%; Pred.No. 2.4e-13;  
Matches 54; Conservative 40; Mismatches 83; Indels 18; Gaps 5;  
QY 33 QFVHLGWVTGGGIGSLKHGNE---IYIAPSGVQKERIQEDMFVCDINEQDISGPPASK 89  
Db 23 RFASFGMRGTSGNLSVLSRDLRLAVTASGHDKGELTPADVLDGDAVAGGRPSA 82  
QY 90 KLKKSQCTPLFMNAYTMRGAGAVITHSKAAVNMATLLFPGQEFKITHQEMIKGIRKCTSG 149  
Db 83 E-----AALHARVAALTGAGAVHVHTVASVANGREPG-GIAFRDLEMLKGV----- 129  
QY 150 GYRYDDMLVVPPIENTPEEKDKERMAHAMNEYDSCAVLVRHGVYVWGETWEKAKTM 209  
Db 130 GLPAHDVEVTLPPVANSQDMKVLGDRLEARD--PRMPAVVAVAGHGLYVWGENPROARHH 187  
QY 210 CECYDYLFDIAVSMK 224  
Db 188 TEVWEMLLELELTQR 202

Search completed: June 1, 2004, 13:36:59  
Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:20:11 ; Search time 17 Seconds  
(without alignments)  
738.171 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSCCQAGGCCSRPCGAQDK.....SMKKMGLDPTQLPVGENGIV 241

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	507	38.5	244	1	XJZ4 YEAST
2	188.5	14.3	208	1	O67788 aquifex aeo
3	152	11.5	181	1	YB18 METUA
4	139	10.6	231	1	SGBE_ECOLI
5	132	10.0	242	1	SGRE_MYCPN
6	128.5	9.8	231	1	SGBE_HAEIN
7	128	9.7	229	1	ARAD_BACSU
8	125.5	9.5	228	1	ARAD_BACST
9	124	9.4	231	1	ARAD_BACSD
10	120	9.1	228	1	SGAE_ECOLI
11	113.5	8.6	231	1	ARAD_SALTY
12	110	8.4	216	1	FUCA_HAEIN
13	109.5	8.3	231	1	ARAD_ECOLI
14	108.5	8.2	215	1	FUCA_ECOLI
15	102.5	7.8	210	1	YGBL_HAEIN
16	98	7.4	284	1	PNK_HELPJ
17	97	7.4	706	1	ADGG_HUMAN
18	96	7.3	705	1	ADGG_RAT
19	93	7.1	706	1	ADGG_MOUSE
20	92.5	7.0	2194	1	GLSN_MEDSA
21	91.5	6.9	212	1	YGBL_ECOLI
22	89.5	6.8	304	1	RGGC_VIBCH
23	89	6.8	536	1	FYN_XIPHE
24	88.5	6.7	284	1	PNK_HELPY
25	86.5	6.6	726	1	ADDB_HUMAN
26	86	6.5	352	1	Z185_MOUSE
27	85	6.5	304	1	RDGC_VIBPA
28	85	6.5	859	1	NIA_FICAN
29	83.5	6.3	725	1	ADDB_MOUSE
30	83.5	6.3	725	1	ADDB_RAT
31	82.5	6.3	1729	1	DME_ARATH
32	80.5	6.1	304	1	RDGC_VIBVU
33	80	6.1	535	1	YRK_CHICK

## RESULT 1

ID	YJZ4 YEAST	STANDARD;	PRT;	244 AA.
AC	P47035;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Hypothetical 27.4 kDa protein in MER2-ENAL intergenic region.			
GN	YJR024C OR J1545 OR YJR83.18.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96109930; PubMed=8619316;			
RA	Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytka J.,			
RA	Sulicka J., Herbert C.J.;			
RT	"The sequence of 24.3 kb from chromosome X reveals five complete open			
RT	reading frames, all of which correspond to new genes, and a tandem			
RT	insertion of a Ty1 transposon.";			
RL	Yeast 11:1179-1186(1995).			
RN	[2]			
RP	SEQUENCE OF 44-244 FROM N.A.			
RC	STRAIN=S288c / FY1679;			
RA	de Haan M., Smits P.H.M., Grivell L.A.;			
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
CC	EMBL; Z49524; CAA89549.1; --			
DR	EMBL; X87297; CAA60719.1; --			
DR	EMBL; X87611; CAA60947.1; --			
DR	PIR; S57042; S57042.			
DR	GermOnline; 141861; --			
DR	SGD; S0003785; YJR024C.			
DR	InterPro; IPR001303; Aldolase_II_N.			
DR	Pfam; PF00596; Aldolase_II; 1.			
DR	Hypothetical protein.			
SK	SEQUENCE 244 AA; 27427 MW; C41FAD142D677A00 CRC64;			

Query Match 38.5%; Score 507; DB 1; Length 244;

Best Local Similarity 46.0%; Pred. No. 1.2e-37;

Matches 108; Conservative 36; Mismatches 67; Indels 24; Gaps 7;

QY 14 PCGAQDKHPRFLPELCKQFYHLGWVTGTGGGSLK--HGNEIYIAPSGVQKRIQPED 71

DB 13 PC-----HPANLICTLCKQFFHNWCTGTGGGSIKDPNTNYIYLAISGVQKRMIFED 66

QY 72 MFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVHTHSKAAVNTLLFPQGE 131

```
Db 67 LFVMDAQTEVLRSP--KLYPSACTPLFLACYQKKNAGAIHHSQNAVICSLF-GDE 123
Qy 132 FKITHQEMIKGI-----RKTSGGYRYDDMLVVPPIIENTPEEKDKLKERMAHANE 182
Db 124 FRIANIEQIKAIPSKQVDPVTKKPWALGFF---DTLKIPPIENMAHEDELIDDLHKTFK 180
Qy 183 YPDSCAVLVRRHGVVWGETWEKATMCCEVDYLFDAVSMKMGCLDPTQLPVGE 237
Db 181 YPDTCAVIVRRHGIFVWGETIDKAKIFNEAIDYLMELAIKTQMGF-PPDCGIGE 234

RESULT 2
YU79_AQUAE STANDARD; PRT; 208 AA.
ID YU79_AQUAE
AC O67788;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical aldolase class II protein AQ_1979.
GN AQ_1979.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RL Nature 392:353-358(1998).
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- SIMILARITY: Belongs to the aldolase class II family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000766; AAC07751.1; -
CC PIR; H70469; H70469.
CC InterPro; IPR001303; Aldolase II_N.
CC DR InterPro; IPR004782; Fuca.
CC DR TIGR; TIGR01086; Fuca; 1.
CC DR TIGR; TIGR01086; Fuca; 1.
CC KW Hypothetical protein; Zinc; Complete proteome.
CC FT METAL 80 80 ZINC (BY SIMILARITY).
CC FT METAL 99 99 ZINC (BY SIMILARITY).
CC FT METAL 101 101 ZINC (BY SIMILARITY).
CC FT METAL 174 174 ZINC (BY SIMILARITY).
CC FT METAL 174 174 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 208 AA; 23550 MW; 3295652C5ED17344 CRC64;

Query Match 14.3%; Score 188.5; DB 1; Length 208;
Best Local Similarity 24.6%; Pred. No. 1.4e-09;
Matches 52; Conservative 39; Mismatches 93; Indels 27; Gaps 5;

Qy 29 ELCKQF-----YHLGWVTGGGSLKHGNE-IYIAPSGVOKERIQPDMEF 73
Db 4 ELFKFKSEKVEEIIIEAGRIILHSRGVWPATSGNISAKVSEEVIAITAGKHKGLTPEIIL 63
Qy 74 VCDINEQDI-SGPPASKKLKKSQCTPLFMNATYMRGAGAVIHTSKAAVMATLLFPQGEF 132
Db 64 LIDIEGRPVGGKGSATLLHTTVYKLFPE-----VNAVVHTSPNATVISIVEKKDFV 117
Qy 133 KITHQEMIKGIRKCTSGGYRYDDMLVVPPIIENTPEEKDKLKERMAHANEYPDSCAVILVR 192
Db 118 ELEDYELLKAF-----PDITHVKIKIPFPNEQNIPLLAKEVENYFKTSEDYKGLIR 172
```

```
Qy 193 RHGVVWGETWEKATMCCEVDYLFDAVSM 223
Db 173 GHGLYTWGRSMEEALIHTEALEFIFECELKL 203

RESULT 3
YE18_METJA STANDARD; PRT; 181 AA.
ID YE18_METJA
AC Q58813;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical aldolase class II protein MJI418.
GN MJI418.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.D., Sadow P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.F., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- SIMILARITY: Belongs to the aldolase class II family. Arab/fuca
CC subfamily.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67582; AAB99428.1; -
CC PIR; A64477; A64477.
CC HSSP; P11550; LDZX.
CC TIGR; MJI418; -
CC InterPro; IPR001303; Aldolase II_N.
CC DR InterPro; IPR00596; Aldolase II; 1.
CC DR Pfam; PF00596; Aldolase; Zinc; Complete proteome.
CC KW Hypothetical protein; Zinc; Complete proteome.
CC FT METAL 68 68 ZINC (BY SIMILARITY).
CC FT METAL 87 87 ZINC (BY SIMILARITY).
CC FT METAL 89 89 ZINC (BY SIMILARITY).
CC FT METAL 147 147 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 181 AA; 20470 MW; E5F3BF13722145B0 CRC64;

Query Match 11.5%; Score 152; DB 1; Length 181;
Best Local Similarity 24.1%; Pred. No. 2e-06;
Matches 42; Conservative 38; Mismatches 66; Indels 28; Gaps 5;

Qy 29 ELCKQFHYLGVWVTGGGSLKHGNEIYIAPSGVOKERIQPDMEFVCDINEQDISPPAS 88
Db 8 KICRKLKYVGVVSGNVSKVKGDKIYLTPTGSLGLFKLEDIDAEWDLGNNVIRKPKTS 67
Qy 89 KLUKKSQCTPLFMNATYMRG-AGAVIHTSKAAVMATLLFPQGEFKITHQEMIKGRKCT 147
Db 68 EK-----NLHLMYRKGRNDINAIHTHSLISTFLSTI--NKEIILLTEGKIFLKKI- 117
Qy 148 SGGYRYDDMLVVPPIIENTPEEKDKLKERMAHANEYPDSCAVILVRRHGVVWGE 201
```

Db 118 --GYVDYVEAGSLKAEETAKR-----DEDVILKNHGVVCLGK 154

## RESULT 4

```
SGBE_ECOLI          STANDARD;          PRT;    231 AA.
ID  SGBE_ECOLI
AC  P37680;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-OCT-1994 (Rel. 30, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable sugar isomerase sgbe (EC 5.1.-.-).
GN  SGBE OR B3583.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=K12 / MG1655;
RX  MEDLINE=94316500; PubMed=8041620;
RA  Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT  "Analysis of the Escherichia coli genome. V. DNA sequence of the
RL  region from 76.0 to 81.5 minutes.";
RL  Nucleic Acids Res. 22:2576-2586(1994).
RN  [2]
RP  DISCUSSION OF SEQUENCE.
RA  Raizer J., Charbit A., Reizer A., Saier M.H. Jr.;
RT  "Novel phosphotransferases system genes revealed by bacterial genome
RT  analysis: operons encoding homologues of sugar-specific permease
RT  domains of the phosphotransferase system and pentose catabolic
RT  enzymes.";
RL  Genome Sci. Technol. 1:53-75(1996).
CC  -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE.
CC  -1- COPACTOR: Binds 1 zinc ion per molecule (Potential).
CC  -1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC  subfamily.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; U00039; AAB19560.1; -.
CC  EMBL; AE000435; AAC76607.1; -.
CC  PIR; S47804; S47804.
CC  EcoGene; EG12287; sgbe.
CC  InterPro; IPR001303; Aldolase II_N.
CC  Pfam; PF00596; Aldolase II; 1.
CC  Isomerase; Zinc; Complete proteome.
KW  METAL 76 76 ZINC (BY SIMILARITY).
FT  METAL 95 95 ZINC (BY SIMILARITY).
FT  METAL 97 97 ZINC (BY SIMILARITY).
FT  METAL 171 171 ZINC (BY SIMILARITY).
SQ  SEQUENCE 231 AA; 25561 MW; PAFRAD7EC2A80B3A CRC64;
Query Match 10.6%; Score 139; DB 1; Length 231;
Best Local Similarity 27.8%; Pred. No. 3.8e-05;
Matches 65; Conservative 27; Mismatches 96; Indels 46; Gaps 13;
QY 26 LIPELCKQFVHLGWTGTGGGSLKHGNE--IYIAPSGVKOKERTOPEDMFVCDINEQDIS 83
DB 10 LAANLALPAHHL--VTFWGNVSVADETRQWVTKPSGVEYDVMTADMVVEI-----AS 63
QY 84 GPPASKKLKKSQCCTPLFWNAYTMRG-AGAVITHSKAAVM---ATLLFPQGEFKITHQEM 139
DB 64 GKVVESGKKSPSDPTHLALRYRAEIGGIIVHTRHATWSQAGLDLPA--WGTTHADY 121
QY 140 IKGIRKCTS-----GGYRYDDMLVPIIENTPEEKOLKERMAMHANEYPDSCAVLVR 192
DB 122 FYGAIPCTQMTABEINGEYEQ---TGEVLIETFEER-----GRSPAQIP---AVLVH 169
```

QY 193 RHGVTVWG-----ETWEKAKTMCEC-YDYLFDAVSMKKMGGLDFTQLPVGENGIV 241  
DB 170 SHGPFAWGKGNRAADAVHNAVLEECAYMGLFSRQLA-----POLPAMQNEILL 215

## RESULT 5

```
SGAE_MYCPN          STANDARD;          PRT;    242 AA.
ID  SGAE_MYCPN
AC  P75289;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable sugar isomerase sgae (EC 5.1.-.-).
GN  SGAE OR MPN498 OR MP345.
OS  Mycoplasma pneumoniae.
OC  Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX  NCBI_TaxID=2104;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=ATCC 29342 / M129;
RX  MEDLINE=97105885; PubMed=8948633;
RA  Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA  Herrmann R.;
RT  "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL  pneumoniae.";
RL  Nucleic Acids Res. 24:4420-4449(1996).
CC  -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE. PROBABLY
CC  INVOLVED IN A METABOLIC PATHWAY WITH SGAH AND SGAU.
CC  -1- COPACTOR: Binds 1 zinc ion per molecule (Potential).
CC  -1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC  subfamily.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; AE000032; AAB95992.1; -.
CC  PIR; S73671; S73671.
CC  InterPro; IPR001303; Aldolase II_N.
CC  Pfam; PF00596; Aldolase II; 1.
CC  Isomerase; Zinc; Complete proteome.
KW  METAL 80 80 ZINC (BY SIMILARITY).
FT  METAL 99 99 ZINC (BY SIMILARITY).
FT  METAL 101 101 ZINC (BY SIMILARITY).
FT  METAL 175 175 ZINC (BY SIMILARITY).
SQ  SEQUENCE 242 AA; 27092 MW; 3F326B0F364CB88B CRC64;
Query Match 10.0%; Score 132; DB 1; Length 242;
Best Local Similarity 27.9%; Pred. No. 0.00017;
Matches 55; Conservative 25; Mismatches 73; Indels 44; Gaps 10;
QY 26 LIPELCKQFVH-----LGWVTGTGGGSLKHGNEIY--IAPSGVKOKERTOPEDMFVCD 76
DB 5 MINDLKEQVFTNLLLPKYGLVHTWGNVSMIAPNRQFVFKPSGVSVDKRAQDMVVVD 64
QY 77 I--NEQDISGPPASKKLKKSQCCTPLFWNAYT-MRGAGAVITHSK-AAVMATLLFPQGEF 132
DB 65 LDNNVLDTNG-----LKPSSDPTTHALMYKCPDKAIIVHTSTFTATSPAQDKPIPL 118
QY 133 KITHQEMIKGIRKCTS-----GGYRYDDMLVPIIENTPEEKOLKERMAMHANEYPD 185
DB 119 GTTHDNFFGPICTTALSDEINGAYEHTGLVI-----LEHLKNNQVD 163
QY 186 --SC-AVLNRRHGVVYV 199
DB 164 VNACAILVKEHGSFVW 180
```



## RESULT 6

SGBE HAEIN STANDARD; PRT; 231 AA.

AC P44989;

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable sugar isomerase sgbe (EC 5.1.-.-).

GN SGBE OR H1025.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI\_TaxID=727;

RN [1]\_TaxID=727;

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

RT Rd.";

RL Science 269:496-512(1995).

CC -!- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE.

CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).

CC -!- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca

CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U32783; AAC22685.1; --

DR TIGR: H11025;

DR InterPro: IPR001303; Aldolase II\_N.

DR Pfam: PF00596; Aldolase II; 1.

KW Isomerase; Zinc; Complete proteome.

FT METAL 76 76 ZINC (BY SIMILARITY).

FT METAL 95 95 ZINC (BY SIMILARITY).

FT METAL 97 97 ZINC (BY SIMILARITY).

FT METAL 171 171 ZINC (BY SIMILARITY).

SQ SEQUENCE 231 AA; 25980 MW; 9DEB3485E54B10DC7 CRC64;

Query Match 9.8%; Score 128.5; DB 1; Length 231;

Best Local Similarity 27.8%; Pred. No. 0.00032;

Matches 50; Conservative 23; Mismatches 74; Indels 33; Gaps 9;

QY 35 YHLGWTGCGISL--KHGNEIYAPSGVQKERIQPEDMFVCDINEQDISGPPASKLK 92

DB 19 HHL--VFTWGNVSAIDREKNLV1KPSGVVDVNTENDMVVDL-----FTGNIVEGNKK 72

QY 93 KSQCTPLFNAY--TWRGAGAVLTHSK--AAVNATLLFFGQEPKITHQEMIKGIRCT--- 147

DB 73 PSSDPTFTHLYRQPHIGVHTSRHATIWAQGLDIEVGTTHGDFYFVGTICTRQM 132

QY 148 -----SGGYRYDDMLVPIIENTPEEKDKERMAHAMNEYDSC-AVLVRHGVYVNGE 201

DB 133 TTKETKGNV---ELETGKVIETVLSRGIE-----PDNIPAVLVHSHGPFAMGK 178

RESULT 7

ARAD\_BACSU STANDARD; PRT; 229 AA.

ID ARAD\_BACSU

P94525; O05186;

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose

DE isomerase).

GN ARAD OR BSU28780.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]\_TaxID=1423;

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=97237725; PubMed=9084180;

RA Sa-Nogueira I.M.G., Nogueira T.V., Soares S., de Lencastre H.;

RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,

RL genetic organization and expression.";

RN Microbiology 143:957-969(1997).

RP [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=97124191; PubMed=8969504;

RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.;

RA Sanders J., Emerson P.T., Harwood C.R.;

RT "The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus

RT subtilis chromosome containing genes responsible for stress

RT responses, the utilization of plant cell walls and primary

RT metabolism.";

RL Microbiology 142:3067-3078(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

RA Guiseppe G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Konigstein G., Krohn S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.";

RL Nature 390:249-256(1997).

RN [4]

RP TRANSCRIPTIONAL REGULATION.

RX MEDLINE=9348379; PubMed=10417639;

RA Mota L.J., Tavares P., Sa-Nogueira I.M.G.;

RT "Mode of action of AraC, the key regulator of L-arabinose metabolism

RT in Bacillus subtilis.";

RL Mol. Microbiol. 33:476-489(1999).

CC -!- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-

CC phosphate.



CC subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF001513; BAB05590.1; -.  
 CC PIR; G83883; G83883.  
 CC InterPro; IPR001303; Aldolase II\_N.  
 CC Pfam; PF00596; Aldolase II; 1.  
 CC Arabinose catabolism; Isomerase; Zinc; Complete proteome.  
 KW METAL 75 75 ZINC (BY SIMILARITY).  
 FT METAL 94 94 ZINC (BY SIMILARITY).  
 FT METAL 96 96 ZINC (BY SIMILARITY).  
 FT METAL 168 168 ZINC (BY SIMILARITY).  
 FT METAL 231 AA; 25991 MW; 6CA570D2A6B4DB97 CRC64;  
 SQ SEQUENCE 231 AA; 25991 MW; 6CA570D2A6B4DB97 CRC64;  
 Query Match 9.4%; Score 124; DB 1; Length 231;  
 Best Local Similarity 28.7%; Pred. No. 0.00081;  
 Matches 51; Conservative 20; Mismatches 67; Indels 40; Gaps 10;  
 CC  
 CC 40 VTGTGGGIS--LKHGNEIYIAPSGVQKRIQPEDMFVCDINEODISGPPASKKLKKSQCT 97  
 CC 22 VFTTGNVSGIDREKGLVVKPSGVYFEMKSKDMVVDLEGNIVG-----DLKPSST 76  
 CC  
 CC 98 PLPMNAY--TMRGAGAVIHTSKAAVMTL-LFPGQE-----FKITHQEMIKGIRKCTS---- 148  
 CC 77 PTHLALYRAPDKVGGIVHTS---VWATAWAQAKGIPAYGTTTHADYFHGTIPCTRPMT 133  
 CC  
 CC 149 -----GGYRYDDMLVPIIETPEEDKERMAMANEYDSC-AVLVRHGVYVNGE 201  
 CC 134 TEILGDYKETGNVIVETFRN---KD-----PMSIPGVLVHSHAFVFWGK 175  
 CC  
 CC RESULT 10  
 CC SGAE\_ECOLI STANDARD; PRT; 228 AA.  
 CC ID SGAE\_ECOLI  
 CC AC P39306;  
 CC DT 01-FEB-1995 (Rel. 31, Created)  
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Probable sugar isomerase sgae (EC 5.1.-.-).  
 CC GN SGAE OR B4198.  
 CC OS Escherichia coli.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC OC Enterobacteriaceae; Escherichia.  
 CC OX NCBI\_TaxID=562;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=K12 / MG1655;  
 CC RX MEDLINE=9534362; PubMed=7610040;  
 CC RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 CC Blattner F.R.;  
 CC RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 CC RT region from 92.8 through 100 minutes."  
 CC RT Nucleic Acids Res. 23:2105-2119(1995).  
 CC RN [2]  
 CC RP DISCUSSION OF SEQUENCE.  
 CC RA Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;  
 CC RT "Novel phosphotransferases system genes revealed by bacterial genome  
 CC RT analysis: operons encoding homologues of sugar-specific permease  
 CC RT domains of the phosphotransferase system and pentose catabolic  
 CC RT enzymes."  
 CC RL Genome Sci. Technol. 1:53-75(1996).  
 CC CC -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE. PROBABLY  
 CC INVOLVED IN A METABOLIC PATHWAY WITH SGAE AND SGAT.  
 CC CC -1- COFACTOR: Binds 1 zinc ion per molecule (Potential).  
 CC CC -1- SIMILARITY: Belongs to the aldolase class II family. AraD/fucA  
 CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U14003; AAA97094.1; -.  
 CC PIR; S56423; S56423.  
 CC EcoGene; EG12498; sgae.  
 CC InterPro; IPR001303; Aldolase II\_N.  
 CC Pfam; PF00596; Aldolase II; 1.  
 CC Isomerase; Zinc; Complete proteome.  
 KW METAL 74 74 ZINC (BY SIMILARITY).  
 FT METAL 93 93 ZINC (BY SIMILARITY).  
 FT METAL 95 95 ZINC (BY SIMILARITY).  
 FT METAL 167 167 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 228 AA; 25278 MW; 3E96E7E0261E36B6 CRC64;  
 Query Match 9.1%; Score 120; DB 1; Length 228;  
 Best Local Similarity 25.8%; Pred. No. 0.0018;  
 Matches 50; Conservative 24; Mismatches 74; Indels 46; Gaps 9;  
 CC  
 CC 21 EHRFLIPELCKQFVHLGWVTGGISL--KHGNEIYIAPSGVQKRIQPEDMFVCDIN 78  
 CC 14 ELPRY-----GLVTTGNVNSAIDRGLWIKPSGVAYETWKAADMVVVDMS 61  
 CC  
 CC 79 EQDISGPPASKKLKKSQCTPLFMNAY--TMRGAGAVIHTSKAA---VMATLLFPFGQEFKI 134  
 CC 62 GKVEG-----EYRSSDTATHELYRRYPSLGGIVHTSTATAWAQAGLAIPA--LCT 114  
 CC  
 CC 135 THQEMIKGIRKCTSG-----GYRYDDMLVPIIETPEEDKERMAMANEYDSC 187  
 CC 115 THADYFFGDIPTCTGLSEEVQGEYELNTGKV--IETLGNAPL-----HTP 160  
 CC  
 CC 188 AVLVRHGVYVNGE 201  
 CC 161 GIVVYQHGPFAWGK 174  
 CC  
 CC RESULT 11  
 CC ARAD\_SALTY STANDARD; PRT; 231 AA.  
 CC ID ARAD\_SALTY  
 CC AC P06130;  
 CC DT 01-JAN-1988 (Rel. 06, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose  
 CC DE isomerase).  
 CC GN ARAD OR STM0101.  
 CC OS Salmonella typhimurium.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC OC Enterobacteriaceae; Salmonella.  
 CC OX NCBI\_TaxID=602;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=LT2;  
 CC RX MEDLINE=85232046; PubMed=3891514;  
 CC RA Lin H.-C., Lei S.-P., Studnicka G., Wilcox G.;  
 CC RT "The arabid operon of Salmonella typhimurium LT2. III. Nucleotide  
 CC RT sequence of arab and its flanking regions, and primary structure of  
 CC RT its product, L-ribulose-5-phosphate 4-epimerase."  
 CC RL Gene 34:129-134(1985).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 CC RX MEDLINE=21534948; PubMed=11677609;  
 CC RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 CC Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 CC Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.; of Salmonella enterica serovar Typhimurium  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2";  
RL Nature 413:852-856 (2001).  
CC -!- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-  
CC phosphate.  
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).  
CC -!- PATHWAY: L-arabinose catabolism; third step. Arad/fuca  
CC -!- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca  
CC subfamily.  
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 202.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC ENBL; M11047; AAA27025.1; ALT FRAME.  
CC EMBL; AE008698; AAL19065.1; -.  
CC PIR; A24986; ISEB4T.  
CC StyGene; SG10015; arad.  
CC InterPro; IPR001303; Aldolase II\_N.  
CC InterPro; IPR004661; Arad.  
CC Pfam; PF00596; Aldolase II; 1.  
CC TIGRFAMs; TIGR00760; arad; 1.  
CC Arabinose catabolism; Isomerase; Zinc; Complete proteome.  
CC METAL 76 76 ZINC (BY SIMILARITY).  
CC METAL 95 95 ZINC (BY SIMILARITY).  
CC METAL 97 97 ZINC (BY SIMILARITY).  
CC METAL 171 171 ZINC (BY SIMILARITY).  
CC SEQUENCE 231 AA; 25531 MW; DA473505739284F9 CRC64;  
Query Match 8.6%; Score 113.5; DB 1; Length 231;  
Best Local Similarity 26.4%; Pred. No. 0.0068;  
Matches 47; Conservative 20; Mismatches 84; Indels 27; Gaps 7;  
QY 40 VTGTGGGSL-KHGNEIYAPSGVKERIQPEDMFVCDINEQDISPPASKLKKSQCT 97  
DB 22 VLTGWNVAVDREGVLVIRPGVDYVMTADNVVSL- - - - -SGEVGHHKPSD 77  
QY 98 PLFMNAY-TMRGAGAVITHSK-AAVMATLPPGQEFKITHQEMIKGRKCTS - - - - - 148  
DB 78 PTHRLLYQAFPTIGGVTHSRHATWAQAGQIPATGTHADYFYGTIPCTRKMTAEAI 137  
QY 149 GGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLRHGVVWGTEWEKA 206  
DB 138 NGEYEW- - - - -TGNVIVTFEKQGI- - - - -DAAQMPGVLVHSHGPFAMGNAEDA 183  
RESULT 12  
FUCA\_HAEIN STANDARD; PRT; 216 AA.  
AC P44777;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE L-fucose phosphate aldolase (EC 4.1.2.17) (L-fucose-1-phosphate  
DE aldolase).  
GN FUCA OR H10611.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxId=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Spriggs T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd";  
RL Science 269:496-512(1995).  
CC -!- CATALYTIC ACTIVITY: L-fucose 1-phosphate = glycerone phosphate +  
CC (S)-lactaldehyde.  
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).  
CC -!- PATHWAY: Fucose metabolism; third step.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca  
CC subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U32743; AAC22270.1; -.  
CC PIR; C64081; C64081.  
CC HSP; P11550; 2FUA.  
CC TIGR; H10611; -.  
CC InterPro; IPR001303; Aldolase II\_N.  
CC InterPro; IPR004782; Fuca.  
CC Pfam; PF00596; Aldolase II; 1.  
CC TIGRFAMs; TIGR01086; fuca; 1.  
CC Lyase; Fucose metabolism; Zinc; Complete proteome.  
CC METAL 73 73 ZINC (BY SIMILARITY).  
CC METAL 92 92 ZINC (BY SIMILARITY).  
CC METAL 94 94 ZINC (BY SIMILARITY).  
CC METAL 155 155 ZINC (BY SIMILARITY).  
CC SEQUENCE 216 AA; 23944 MW; 4DBEE21ED8EA6C6 CRC64;  
Query Match 8.4%; Score 110; DB 1; Length 216;  
Best Local Similarity 23.3%; Pred. No. 0.013;  
Matches 48; Conservative 34; Mismatches 96; Indels 28; Gaps 6;  
QY 27 IPELCKQFHLGWTGTGGGSLKHGNEIYAPSGVKERIQPEDMFVCDINEQDISGPP 86  
DB 10 IIDTCLEMTKLGNGQTAGNVSVRYKDGMLTPTGMPYHLMKNTENIVVDGN- - - - -GKH 64  
QY 87 ASKKLKKSQCTPLFMNAYTMRGAGAVITHSKAAVMATL- - - - -PPGQEFKITHQEMIKGI 143  
DB 65 EENKLPSEWQFHLVSVHTPREANAVVNHSHHCAGLSILEKPIPIAHMV- - - - - 115  
QY 144 RKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLRHGVVWGTEW 203  
DB 116 - - - - -AVSG- - - - -TDHIPCPVYA-TFGSHKLSYVATGIKE- - - - -SKALLAHGLITCGENL 164  
QY 204 EKATMCECYDYLFDIAVSKMKMGILD 229  
DB 165 DKALMLAQEVEVLASVWLKLLSTGLE 190  
RESULT 13  
ARAD\_ECOLI STANDARD; PRT; 231 AA.  
ID ARAD\_ECOLI  
AC P08203;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose  
DE isomerase).  
GN ARAD OR B0061.

```

OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=B;
RX MEDLINE=87163495; PubMed=3549454;
RA Lee N., Gleisow W., Martin R., Hamilton E., Fowler A.;
RT "The organization of the araBAD operon of Escherichia coli.";
RL Gene 47:231-244(1986).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=91083835; PubMed=2261080;
RA Chen H., Sun Y., Stark T., Beattie W., Moses R.E.;
RT "Nucleotide sequence and deletion analysis of the polB gene of
RT Escherichia coli.";
RL DNA Cell Biol. 9:631-635(1990).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=91067495; PubMed=2251150;
RA Mineno J., Fukui H., Ishino Y., Kato I., Shinagawa H.;
RT "Nucleotide sequence of the araD gene of Escherichia coli K12
RT encoding the L-ribulose 5-phosphate 4-epimerase.";
RL Nucleic Acids Res. 18:6722-6722(1990).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [6]
RN SEQUENCE OF 222-231 FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=91017565; PubMed=2217198;
RA Bonner C.A., Hays S., McEntee K., Goodman M.F.;
RT "DNA polymerase II is encoded by the DNA damage-inducible dinA gene
RT of Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7663-7667(1990).
RN [7]
RN SEQUENCE OF 158-231 FROM N.A.
RC STRAIN=K12;
RC STRAIN=K12;
RX MEDLINE=91238699; PubMed=2034216;
RA Iwasaki H., Ishino Y., Toh H., Nakata A., Shinagawa H.;
RT "Escherichia coli DNA polymerase II is homologous to alpha-like DNA
RT polymerases.";
RL Mol. Gen. Genet. 226:24-33(1991).
CC -1- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-
CC phosphate.
CC -1- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -1- PATHWAY: L-arabinose catabolism; third step.
CC -1- SIMILARITY: Belongs to the aldolase class II family. AraD/fucA
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

---

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M15263; AAA23464.1; -; NOT ANNOTATED CDS.
DR EMBL; M35371; -; NOT ANNOTATED CDS.
DR EMBL; M62646; AAA24405.1; -;
DR EMBL; D10483; BAB96630.1; -;
DR EMBL; AE000116; AAC73172.1; -;
DR EMBL; M37727; AAA23683.1; -;
DR EMBL; M38283; AAA63763.1; -;
DR EMBL; X56048; CAA39519.1; -;
DR PIR; E64727; ISECP4.
DR PDB; 1JDI; 23-JAN-02.
DR PDB; 1KOW; 28-JAN-03.
DR ECO2DBASE; G028.1; 6TH EDITION.
DR EcoGene; EG10055; arad.
DR InterPro; IPR001303; Aldolase_II_N.
DR InterPro; IPR004661; Arad.
DR Pfam; PF00596; Aldolase II; 1.
DR TIGRFAMs; TIGR00760; arad; 1.
KW Arabinose catabolism; Isomerase; Zinc; Complete proteome;
KW 3D-structure.
FT METAL 76 76 ZINC (BY SIMILARITY).
FT METAL 95 95 ZINC (BY SIMILARITY).
FT METAL 97 97 ZINC (BY SIMILARITY).
FT METAL 171 171 ZINC (BY SIMILARITY).
FT VARIANT 50 50 V -> I.
FT VARIANT 70 70 T -> A.
FT VARIANT 216 216 D -> N.
SQ SEQUENCE 231 AA; 25519 MW; 1753F75958332163 CRC64;
Query Match 8.3%; Score 109.5; DB 1; Length 231;
Best Local Similarity 26.1%; Pred. No. 0.015;
Matches 47; Conservative 19; Mismatches 83; Indels 31; Gaps 8;
QY 40 VTGTGGGIGSL--KHGNEIYIAPSGVQKRIQEDMFVCDINEQDISGPPASKKLKKSQCT 97
DB 22 VTLTWGNVSAVDRERGVRGKPSGVYDYSVMTADDMVVVSIE---TGEVVEGTKKPSDT 77
QY 98 PLFMNAY--TMRGAGAVIHTHSAKAAVMATLLPQQEFK---ITHQEMIKGIKICTS----- 148
DB 78 PTHRLLYQAPPSIGGIVHSHRHAIWAQ--AGQSIPTGTHADYFYGTICTRKMTDA 135
QY 149 --GGYRYDDMLVPIIENPEEKDKERMAHAMNEYDSCAVLRRHGVVYVWGETWEKA 206
DB 136 EINGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVHSHGPFANGKNAEDA 183

```

---

```

RESULT 14
FUCA_ECOLI STANDARD; PRT; 215 AA.
AC P11550;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE L-fuculose phosphate aldolase (EC 4.1.2.17) (L-fuculose-1-phosphate
DE aldolase).
GN FUCA OR FUCC OR PRD OR B2800 OR SF2814 OR S3009.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12;
RC MEDLINE=89315234; PubMed=2664711;
RX Lu Z., Lin B.C.C.;
RA "The nucleotide sequence of Escherichia coli genes for L-fucose
RT dissimilation";
RL Nucleic Acids Res. 17:4883-4884(1989).
RN [2]

```

RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=90036697; PubMed=2553671;  
RA Chen Y.M., Lu Z., Lin E.C.C.;  
RT "Constitutive activation of the fucAO operon and silencing of the  
RT divergently transcribed fucPK operon by an ISS element in  
RT *Escherichia coli* mutants selected for growth on L-1,2-propanediol.";  
RL J. Bacteriol. 171:6097-6105(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MGL655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE OF 108-215 FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=89291720; PubMed=2661535;  
RA Conway T., Ingram L.O.;  
RT "Similarity of *Escherichia coli* propanediol oxidoreductase (fucO  
RT product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis*  
RT and *Saccharomyces cerevisiae*.";  
RL J. Bacteriol. 171:3754-3759(1989).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=22272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
RT through comparison with genomes of *Escherichia coli* K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.N., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of *Shigella*  
RT flexneri serotype 2a strain 2457T.";  
RL Infect. Immun. 71:2775-2786(2003).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS), AND CATALYTIC MECHANISM.  
RC SPECIES=E.coli;  
RX MEDLINE=96256522; PubMed=8676381;  
RA Dreyer M.K., Schulz G.E.;  
RT "Catalytic mechanism of the metal-dependent fuculose aldolase from  
RT *Escherichia coli* as derived from the structure.";  
RL J. Mol. Biol. 259:458-466(1996).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
RC SPECIES=E.coli;  
RX MEDLINE=96256522; PubMed=8676381;  
RA Dreyer M.K., Schulz G.E.;  
RT "Refined high-resolution structure of the metal-ion dependent L-  
RT fuculose-1-phosphate aldolase (class II) from *Escherichia coli*.";  
RL Acta Crystallogr. D 52:1082-1091(1996).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS), AND MUTAGENESIS.  
RC SPECIES=E.coli;  
RX MEDLINE=20281325; PubMed=10821675;  
RA Joergers A.C., Grosse C., Fessler W.D., Schulz G.E.;  
RT "Catalytic action of fuculose 1-phosphate aldolase (class II) as  
RT derived from structure-directed mutagenesis.";  
RL Biochemistry 39:6033-6041(2000).

RN X-RAY CRYSTALLOGRAPHY OF MUTANTS.  
RP SPECIES=E.coli;  
RX MEDLINE=20510153; PubMed=11054289;  
RA Joergers A.C., Mueller-Dieckmann C., Schulz G.E.;  
RT "Structures of L-fuculose-1-phosphate aldolase mutants outlining  
RT motions during catalysis.";  
RL J. Mol. Biol. 303:531-543(2000).  
RN [10]  
CC - CATALYTIC ACTIVITY: L-fuculose 1-phosphate = glycerone phosphate +  
CC (S)-lactaldehyde.  
CC - COFACTOR: Binds 1 zinc ion per molecule.  
CC - PATHWAY: Fucose metabolism; third step.  
CC - SUBUNIT: Homotetramer.  
CC - SIMILARITY: Belongs to the aldolase class II family. Arab/fuca  
CC subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M31059; AAA23823.1; -;  
CC EMBL; X15025; CAA33125.1; -;  
CC EMBL; U29581; AAB40450.1; -;  
CC EMBL; AE000363; AAC75842.1; -;  
CC EMBL; M27177; -; NOT ANNOTATED\_CDS.  
CC EMBL; AE015295; AAN44302.1; -;  
CC EMBL; AE016987; AAP18127.1; -;  
CC PIR; B33495; ADECFP.  
CC PDB; 1FUA; 14-OCT-96.  
CC PDB; 2FUA; 14-OCT-96.  
CC PDB; 3FUA; 14-OCT-96.  
CC PDB; 4FUA; 14-OCT-96.  
CC PDB; 1E46; 18-JUL-03.  
CC PDB; 1E47; 18-JUL-03.  
CC PDB; 1E48; 18-JUL-03.  
CC PDB; 1E49; 18-JUL-03.  
CC PDB; 1E4A; 18-JUL-03.  
CC PDB; 1E4B; 18-JUL-03.  
CC PDB; 1E4C; 18-JUL-03.  
CC PDB; 1D2V; 21-FEB-02.  
CC PDB; 1D2U; 20-JUL-00.  
CC PDB; 1D2W; 21-FEB-02.  
CC PDB; 1D2X; 21-FEB-02.  
CC PDB; 1D2Y; 24-JUN-03.  
CC PDB; 1D2Z; 21-FEB-02.  
CC ECoGene; EGI0348; fuca.  
CC InterPro; IPR001303; Aldolase\_II\_N.  
CC Pfam; PF00596; Aldolase\_II; 1.  
CC TIGRPFAM; TIGR01086; fuca; 1.  
KW Lyase; Fucose metabolism; Zinc; 3D-structure; Complete proteome.  
FT METAL 73 73 ZINC.  
FT METAL 92 92 ZINC.  
FT METAL 94 94 ZINC.  
FT METAL 155 155 ZINC.  
FT HELIX 3 19  
FT TURN 20 21  
FT HELIX 24 26  
FT STRAND 29 34  
FT TURN 35 36  
FT STRAND 37 40  
FT STRAND 42 42  
FT HELIX 47 49  
FT HELIX 52 54  
FT STRAND 56 58  
FT TURN 60 61  
FT STRAND 64 64  
FT TURN 66 67  
FT STRAND 70 70

FT TURN 72 73  
FT HELIX 74 83  
FT TURN 85 86  
FT STRAND 89 93  
FT TURN 96 104  
FT TURN 105 105  
FT STRAND 108 108  
FT HELIX 113 118  
FT TURN 119 119  
FT STRAND 123 123  
FT STRAND 125 126  
FT TURN 131 132  
FT TURN 134 143  
FT STRAND 149 152  
FT TURN 153 155  
FT STRAND 156 161  
FT HELIX 164 185  
FT TURN 186 187  
FT HELIX 196 209  
SQ SEQUENCE 215 AA; 23775 MW; BA9897E13ABE4A22 CRC64;

Query Match 8.2%; Score 108.5; DB 1; Length 215;  
Best Local Similarity 23.9%; Pred. No. 0.017;  
Matches 52; Conservative 37; Mismatches 86; Indels 43; Gaps 11;

QY 27 IPELCKQFYHLGWVTGGTGGISLKHGNEIYIAPSGVKERIQPDMFVCDINEQDISGPP 86  
DB 10 IDTCTLEWRLGNQGTAGNSVRYQDGMILPTGIPYEKLTHESHVFDGNGKHEEG-- 67  
QY 87 ASKKUKKSQCTPLFNWYTMR-GAGAVITHSKAAVMATLL---PPQGEFKITHQEMIKG 142  
DB 68 ---KLPSSEWR-FHMAAYQSRPDANAVVHNHVAHVCTAVSILNRSIPAIHYMI----- 115  
QY 143 IRKTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLVRRHGVVWGFT 202  
DB 116 ---AAAGG---NSIPCAVYA-TFTRELSERVALALK---NRKATLQHGHLIACEVN 163  
QY 203 WEKAKTMCBCYD-----YLFDAVSMKMGMLDPTPLPV 235  
DB 164 LEKALWLAHEVEVLAQLYLTTLAIT-----DP--VPV 193

RESULT 15  
YGBL HAEIN  
ID YGBL HAEIN STANDARD; PRT; 210 AA.  
AC Q57199; O05040;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical aldolase class II protein H11012.  
GN H11012.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.I., Geoghagen N.S.M.,  
RA Gnehm C.I., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -I- COFACTOR: Binds 1 zinc ion per molecule (Potential).  
CC -I- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca

subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32782; AAC22673.1; -;  
DR PIR; B64108; B64108.  
DR TIGR; H11012; -;  
DR InterPro; IPR001303; Aldolase II\_N.  
DR Pfam; PF00596; Aldolase II; 1.  
KW Hypothetical protein; zinc; Complete proteome.  
FT METAL 74 74 ZINC (BY SIMILARITY).  
FT METAL 93 93 ZINC (BY SIMILARITY).  
FT METAL 95 95 ZINC (BY SIMILARITY).  
FT METAL 160 160 ZINC (BY SIMILARITY).  
SQ SEQUENCE 210 AA; 23242 MW; 4800CA987DBDF01F CRC64;

Query Match 7.8%; Score 102.5; DB 1; Length 210;  
Best Local Similarity 22.3%; Pred. No. 0.058;  
Matches 41; Conservative 26; Mismatches 82; Indels 35; Gaps 6;

QY 26 LIPELCKQFYHLGWVTGGTGGISLKHGNEIYIAPSGVKERIQPDMFVCDINEQDISG 84  
DB 9 LMVQLGRSFYERGIYVGGAGNLSVLDNRVLVPTGSSLGRLSVRLSVLDMEGNLLGG 68  
QY 85 PPASKKL-----KKSQCTPLFNWYTMRGAGAVITHSKAAVMATLLPPQGEFKITH 136  
DB 69 DKPSKEAVFHLAMYKKNPECK-----AIVHLHSTYLTALSCL-----DN 107  
QY 137 QEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLVRRHGV 196  
DB 108 LDPNVAIPPTFYVMRVGKQVLIYYR--PGSPKIAELS---NRALTGKAFLLAHGV 162  
QY 197 YVWG 200  
DB 163 VTG 166

Search completed: June 1, 2004, 13:33:52  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:30:06 ; Search time 20 seconds  
(without alignments)  
1159.108 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCRCRPGCAQDK.....SMKMGELDPTQLPVGNGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	38.5	244	2	S57042
2	433.5	32.9	284	2	T27523
3	274.5	20.8	192	2	T39191
4	219.5	16.7	209	2	A69864
5	188.5	14.3	208	1	H70469
6	164	12.5	205	2	D83436
7	153	11.6	218	2	B82587
8	152	11.5	181	1	A64477
9	148.5	11.3	254	1	A72396
10	140	10.6	244	2	AC3533
11	139	10.6	231	2	S47804
12	138.5	10.5	191	1	C69054
13	132	10.0	231	2	AI0977
14	132	10.0	242	2	S73671
15	132	10.0	243	2	G90586
16	130	9.9	227	2	A95238
17	128.5	9.8	243	2	H64108
18	128	9.7	229	2	E59587
19	127	9.6	211	2	H84275
20	126	9.6	230	2	B42484
21	125	9.5	234	2	B98102
22	124	9.4	231	2	G83883
23	123.5	9.4	215	1	E72546
24	122.5	9.3	231	2	AI0271
25	122	9.3	228	2	F91275
26	122	9.3	228	2	F86116
27	121	9.2	228	2	AD1052
28	120	9.1	228	2	S56423
29	115.5	8.8	231	2	AB0404

30	114	8.7	189	1	E71241	L-fuculose-phospha
31	113.5	8.6	248	1	ISEB4T	L-ribulose-phospha
32	111.5	8.5	212	2	C95253	L-fuculose-phospha
33	111.5	8.5	215	2	AH0862	fuculose-1-phospha
34	111.5	8.5	217	2	A98118	L-fuculose-phospha
35	110.5	8.4	231	2	AB0515	L-ribulose-5-phosp
36	110.5	8.4	233	2	B97065	ribulose-5-phospha
37	110	8.4	216	1	C64081	L-fuculose-phospha
38	109.5	8.3	231	1	ISECP4	L-ribulose-phospha
39	109.5	8.3	231	2	A90637	L-ribulose-5-phosp
40	109.5	8.3	231	2	A85488	L-ribulose-5-phosp
41	108.5	8.2	215	1	ADECFP	L-fuculose-phospha
42	108.5	8.2	215	2	D91086	L-fuculose-1-phosp
43	108.5	8.2	215	2	F85931	L-fuculose-1-phosp
44	102.5	7.8	210	1	B64108	L-fuculose-phospha
45	99.5	7.6	224	2	AF3646	1-fuculose-phospha

ALIGNMENTS

RESULT 1

S57042  
hypothetical protein YUR024c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein J1545; hypothetical protein YUR83.18  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jul-1995 #sequence revision 08-Sep-1995 #text\_change 19-Apr-2002  
C;Accession: S57042; S57039; S55213; S60503; S61120  
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber  
submitted to the Protein Sequence Database, September 1995  
A;Reference number: S57040  
A;Accession: S57042  
A;Molecule type: DNA  
A;Residues: 1-244 <ZAG>  
A;Cross-references: EMBL:Z49524; NID:G1015662; PIDN:CAA89549.1; PID:G1015663; MIPS:YUR024  
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.  
submitted to the Protein Sequence Database, September 1995  
A;Reference number: S56771  
A;Accession: S57039  
A;Molecule type: DNA  
A;Residues: 44-244 <ZAG>  
A;Cross-references: EMBL:Z49524; MIPS:YUR024c  
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S55183  
A;Accession: S55213  
A;Molecule type: DNA  
A;Residues: 44-244 <DEH>  
A;Cross-references: EMBL:X87611; NID:G854567; PIDN:CAA60947.1; PID:G854598  
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber  
Yeast 11, 1179-1186, 1995  
A;Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fr  
A;Reference number: S60503; MUID:96109930; PMID:8619316  
A;Accession: S60503  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-244 <ZAF>  
A;Cross-references: EMBL:X87297; NID:G1129159; PIDN:CAA60719.1; PID:e183285; PID:G112916  
C;Genetics:  
A;Cross-references: SGD:S0003785  
A;Map position: 10R

Query Match 38.5%; Score 507; DB 2; Length 244;

Best Local Similarity 46.0%; Pred. No. 4, 7e-38;  
Matches 108; Conservative 36; Mismatches 67; Indels 24; Gaps 7;

Qy 14 PCGAQDKHEPRFLPELCKQFYHLGWVTGTGGGSLK--HGNEIYIAPSGVQKERIOPED 71

Db 13 PC-----HPANLICTLCKQFFHNWCTGTGGGSIKDPNTNYIYIAPSGVQKEKMPED 66

Qy 72 MFVCDINEQDISPPASKKLKKSQCTELFNWYTMRCGAVIHTHSAAYWATLLFPQGE 131

Db 67 LFVMDAQTLEYLRSP--KLYKPSACTFLACYQKQKAGAIHTHSONAVICSLLF-GDE 123



QY 132 FKITHQEMIKGI-----RKCTSGGYRYDDMLVPIIENPEEKDLKERMAHMANNE 182  
 DB 124 FRIANIEQIKALPSGKVDPTVKPMALSPF---DTLKPIIENMAHEDDLIDHLHKTFKD 180  
 QY 183 YPDSCAVLRHGVVYVWGVTWEKAKTMCECYDYLFDIAVSMKMGIDPPTQLPVG 237  
 DB 181 YPDTCAVLRHGVVYVWGVTIDKAKIFNEAIDYLMELAKYQMGII-PDCCGIGE 234

## RESULT 2

T27523

hypothetical protein ZC373.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T27523

R:Kershaw, J.

submitted to the EMBL Data Library, April 1995

A:Reference number: Z20382

A:Accession: T27523

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-284 &lt;WIL&gt;

A:Cross-references: EMBL:Z49131; PIDN:CAA88977.1; GSPDB:GN00028; CBSP:ZC373.5

A:Experimental source: clone ZC373

C:Genetics:

A:Gene: CBSP:ZC373.5

A:Map position: X

A:Introns: 26/3; 75/2; 114/1; 236/2

Query Match

Best Local Similarity 32.9%; Score 433.5; DB 2; Length 284;

Matches 87; Conservative 35; Mismatches 75; Indels 21; Gaps 1;

QY

29 ELCKQFYHLGWVTGTGGISLKHGNEIYIAPSGVQKERIQPEDMFVCDINEQDISGPPAS 88

DB

35 ELMTQFYKLGMRGSGGAMGCSISSELMISPSALQKERIREQDVVYVNMKDKTEVQRPNN 94

QY

89 KLUKKSQCTPLFMNAYTMRGAGAVIHTHSAANVATLLFPQGEFKITHQEMIKGIRKCTS 148

DB

95 KRITVSCSVLFSLLMKETGSECVIHTHSCANLITQLIKSNVFEISHQYIIRGIVDPFS 154

QY

149 GGYRYDDMLVPIIENPEEKDL-----KERMAHMANNEVPDSC 187

DB

155 GKALKYSDTITPIDNMPSESQLLVCANRSPMQPCFNIALFQPIRGVLENYPOAI 214

QY

188 AVLVRHGVVYVWGVTWEKAKTMCECYDYLFDIAVSMKK 225

DB

215 AVLVRHGLFWGPTWSTKIMTECIDYLLLESLIEMLK 252

## RESULT 3

T39191

conserved hypothetical protein SPAC9.06c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C&gt;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T39191

R:Wedler, H.; Duesterhoeft, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1999

A:Reference number: Z21834

A:Accession: T39191

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-192 &lt;WED&gt;

A:Cross-references: EMBL:AL121764; PIDN:CAB57424.1; GSPDB:GN00066; SPDB:SPAC9.06c

A:Experimental source: strain 972h; cosmid c9

C:Genetics:

A:Gene: SPDB:SPAC9.06c

A:Map position: 1

A:Introns: 27/2; 53/2; 73/2; 129/1

Query Match

Best Local Similarity 20.8%; Score 274.5; DB 2; Length 192;

Matches 34.6%; Pred. No. 2.9e-17;

Matches 71; Conservative 36; Mismatches 73; Indels 25; Gaps 7;

QY 26 LIPELCKQFYHLGWVT-GTGGISLKHGNEIYIAPSGVQKERIQPEDMFVCDINEQDISG 84

DB 9 LILELIPHYISLGNWKFSGYAIQVK-----DRVQDFITENDIVTFNLSNQSVT- 58

QY 85 PPASKKLKKSQCTPLFMNAYTMRGAGAV--IHTHSAANVATLLFPQGEFKITHQEMIKG 142

DB 59 -----KDLVWAYIFSWVLSNMDAVACIVSYSTVAAGASMY--NEKFTTQSEMIKG 108

QY 143 IRKCT-SGGYRYDDMLVPIIENPEEKDLKERMAHMANNEVPDSCAVLRHGVVYVWGE 201

DB 109 IPKGNPSAGYLCCFDTLEVPIIHN-GDSKTILDELKVIELYPOTCAVLRHGVIGWGA 167

QY 202 TWEKAKTMCECYDYLFDIAVSMKKM 226

DB 168 TWEKSKTQMECYEYLFELDYLKLT 192

## RESULT 4

A69864

conserved hypothetical protein ykry - Bacillus subtilis

C:Species: Bacillus subtilis

C&gt;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: A69864

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Galizzi, A.; Gallier

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Togmon, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: A69864

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-209 &lt;KUN&gt;

A:Cross-references: GB:Z99111; GB:AL009126; NID:G2633699; PIDN:CAB13234.1; PID:G2633732

A:Experimental source: strain 168

C:Genetics:

A:Gene: ykry

C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match

Best Local Similarity 16.7%; Score 219.5; DB 2; Length 209;

Matches 64; Conservative 34; Mismatches 88; Indels 21; Gaps 7;

QY

27 IPBLCKQFYHLGWVTGTGGISLKHGNE---IYIAPSGVQKERIQPEDMFVCDINEODIS 83

DB

11 LAEVKRELAERDFPATSGNLSIKVTDEPLTFLVTASGDKKRBETVEDFLLDQN----- 65

QY

84 GPPASK--KLKKSQCTPLFMNAYTMRGAGAVIHTHSAANVATLLFPQGEFKITH--QEM 139

DB

66 GEPASGHSLSKPSAETLLTHLYNKTNAGCCLHVHTVNNVISELY-GDQKKTFFKQEI 124

QY

140 IKGIRKCTSGGYRYDDMLVPIIENPEEKDLKERMAHMANNEVPDSCAVLRHGVVYVW 199

DB

125 IKAL-----GLWEENAETVPIIENPAHIPTLALFAEISE--DSGAVLRHNGITAW 176

QY

200 GETWEKAKTMCECYDYLFDIAVSMKKM 226

DB

177 GKTAFAKRVLEAVEFLFSYHLKLT 203

## RESULT 5

H70469

L-fuculose-phosphate aldolase homolog - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: H70469  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, D.

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: H70469  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-208 <AQF>  
A:Cross-references: GB:AE000766; GB:AE000657; NID:g2984216; PIDN:AAC07751.1; PID:g298422

A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: fucA2  
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 14.3%; Score 188.5; DB 1; Length 208;  
Best Local Similarity 24.6%; Pred. No. 1.7e-09;  
Matches 52; Conservative 39; Mismatches 93; Indels 27; Gaps 5;

QY 29 ELCKQF-----YHLGWTGTGGGSLKHGNE-IYIAPSGVQKRIQPEDMF 73

Db 4 ELFKFSEKVEEIEAGRIILHSRGWVPATSGNISAKVSEYIATASGKHGKLTPEIDIL 63

QY 74 VDINEQDI-SGPPASKLKKSOCTPLFNAYTMRGAGAVIHTHSAKAAVWATLLFPQGF 132

Db 64 LIDYGRPGGKPSAETLLHTTVYKLFPE-----VNAVHTHSPNATVISIVEKKDFV 117

QY 133 KTHQEMIGIRKCTSGGYRYDDMLVVPPIENTPEKDKERMAHAMNEYPDSCAVLVR 192

Db 118 ELEDYELLKAF-----PDITHTEVKIKIPFPNEQNIPLAKEVNYFTSEDKYGLIR 172

QY 193 RHGVYVWGETWEKAKTMCCECYDLYFDIAYSM 223

Db 173 GHGLYWGSRMEEALHTHEALEFIFFCELKL 203

RESULT 6

D83436  
probable sugar aldolase PA1683 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83436

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brinkman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83436

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-205 <STO>

A:Cross-references: GB:AE004595; GB:AE004091; NID:g9947644; PIDN:AAG05072.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1683

Query Match 12.5%; Score 164; DB 2; Length 205;  
Best Local Similarity 25.9%; Pred. No. 2.7e-07;  
Matches 51; Conservative 28; Mismatches 98; Indels 20; Gaps 5;

QY 24 RFLPELCKQFYHLGWTGTGGGSLKHGNE-IYIAPSGVQKRIQPEDMFVCDINEQDI 82

Db 17 RFL-----YRGWSPATSSNYSARLDQRALLTVSGKHGKGLGDDVLA-----TDL 63

QY 83 SGPPASKLKKSOCTPLFNAYTMRGAGAVIHTHSAKAAVWATLLFPQGFQFKITHQEMIK 141

Db 64 AGNSLEPGKPSAETLLHTTQYANPAIGAVALHTHSAKAAVWATLLFPQGFQFKITHQEMIK 123

QY 142 GIRKCTSGGYRYDDMLVVPPIENTPEKDKERMAHAMNEYPDSCAVLVRHGHVYVWGE 201  
Db 124 AF-----AGVTTHGQVEVPPIFDNDQDIARLASRVQVPLEAHPHCPGGLIRGHGLYTGA 178

QY 202 TWEKAKTMCCECYDLYFD 218

Db 179 RMSDALRQVEAFELFE 195

RESULT 7

E82587

conserved hypothetical protein XP2209 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: E82587

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82587

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-218 <SM>

A:Cross-references: GB:AE004033; GB:AE003849; NID:g9107342; PIDN:AAF85008.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.

as-Neco, E.; Docena, C.; El-Dorri, H.; Pacincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Mitacca, R.C.; Miyaki, C.Y.

; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.C.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XP2209

Query Match 11.6%; Score 153; DB 2; Length 218;

Best Local Similarity 23.2%; Pred. No. 2.8e-06;

Matches 48; Conservative 34; Mismatches 91; Indels 34; Gaps 7;

QY 33 QFVHLGWVTGGGSLK-HGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGP--PASK 89

Db 29 EFAQAGTTPATSSNFSHRLDEHVAITVSGRDKRCLTEEDIMAVDLGNVAGHPHTPSAE 88

QY 90 KKKKSCTPLEFNAYTMRGAGAVIHTHSAKAAVWATLLFPQO-EFKITHQEMIKGRKCTS 148

Db 89 TLLHTLYRRFPE-----IGCVLHTSHLTVTVASRYAGAGHISKDYELLKAFE---- 138

QY 149 GGYRYDDMLVVPPIENTPEKDL-----KERMAHAMNEYPDSCAVLVRHGHVYVW 199

Db 139 -GSHTEHTILDVFPVFCNTQNNILAAQVDTLLDKQRM-----WGLINGHGMVYTW 187

QY 200 GETWEKAKTMCCECYDLYFDIAYSMKQM 226

Db 188 GNTLADARRHLEALEFLLHCELNLKL 214

RESULT 8

A64477

L-fuculose-phosphate aldolase homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000

C:Accession: A64477

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kane, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C.  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64477  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-181 <BUL>  
A:Cross-references: GB:U67582; GB:I77117; NID:gl592064; PIDN:AAB99428.1; PID:gl592067; T  
C:Genetics:  
A:Map position: FOR1381152-1381697  
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 11.5%; Score 152; DB 1; Length 181;  
Best Local Similarity 24.1%; Pred. No. 2.8e-06;  
Matches 42; Conservative 38; Mismatches 66; Indels 28; Gaps 5;

QY 29 ELCKQFHLGWVTGTGGGSLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPAS 88  
DB 8 KICKRLYDRKVVGGGNSVKEGKIYLTPTGILGFLKEDDAEMDLGDNVIGKPTS 67  
QY 89 KKLKKSQCTPLFMNAYTWRG-AGAVIHTHSCAAVWATLLFPQGFKITHQEMIKIRKCT 147  
DB 68 EK-----NLHLMYRKNDINAIHTHSLISTFLSTI--NKBIELITPBGKIFLKKI- 117  
QY 148 SGGYRYDDMLVPIENTPEKDKERMAHMEYDPCAVLRHGGVYVWGE 201  
DB 118 --GYVDYEEAGSLAEETAKR-----DEDVILLKNGHGVCLGK 154

RESULT 9  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: A72396  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <ARN>  
A:Cross-references: GB:AE001710; GB:AE000512; NID:g4980775; PIDN:AAD35371.1; PID:g498078  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TMO283  
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 11.3%; Score 148.5; DB 1; Length 254;  
Best Local Similarity 26.2%; Pred. No. 8.6e-06;  
Matches 64; Conservative 44; Mismatches 95; Indels 41; Gaps 13;

QY 4 COAQGCCSRPCGAQ---DKHPRFLIPELCKQFHLGWVTGTGGGSLKHGNEIYIAPS 60  
DB 27 CVDNEDCTFTHGKGSMEYKERKLYNAHLLEKY--GLVAYTSGNVSVRIGDHVLIKPS 84  
QY 61 GVQKRIQPEDMFVCDINEQDISGPPASKKLSQCTPLFMNAYTWRGAGAVIHTHSCAA 120  
DB 85 GVPYTELKEDFVVDLEGNVIEG---EKKPSVDATYLYLYKH-LDWAKSVIHTHTIFA 140  
QY 121 VMATLLFPQGFKITHQEMIKIRKCTSGYRYDDML--VVPITENTPEEKD-LKERN 177  
DB 141 MWAIL-----EKSIPVL--CTA-----HADVFGEIPLTEYAPVGSEAIKAVV 183  
QY 178 HAMNEYDPCAVLRHGGVYVWGETWEKA--KTMCEYDYLFDIAVS--MKNGLDPTOL 233  
DB 184 KVIGK---SGAVLLRKHGVMVGTSDVDVAKAI-----FLEEVAKAAYFATLAKPTPL 235

QY 234 PVGE 237  
DB 236 PPDE 239

RESULT 10  
A:Title: fuculose phosphate aldolase (EC 4.1.2.17) [imported] - Brucella melitensis (strain 16M)  
A:Reference number: A3533  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-244 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL53430.1; PID:gl7984327; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME110189  
A:Map position: II  
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 10.6%; Score 140; DB 2; Length 244;  
Best Local Similarity 23.0%; Pred. No. 4.8e-05;  
Matches 52; Conservative 33; Mismatches 75; Indels 66; Gaps 10;

QY 4 COAQGCCSRPCGAQDKHPRFLIPELCKQFHLGWVTGTGGGSLKHGNEIYIAPSGVQ 63  
DB 24 CKQMNCDSALLARQS-----IVDAMRSGPEKGFHGGSGNISVREGGHIWVTPGA- 75  
QY 64 KERIQEDMFVCDINEQDISGPPASKKLSQCTPLFMNAYTWRG---AGAVIHTHSCAA 120  
DB 76 TSTWDQDMSLVSLGEGHLAGKIPSEWR-----IHTIMEAHPEAGAVVISHADAC 127  
QY 121 VMATLL-----FPQGFKITHQEMIKIRKCTSGYRY--DDMLVVPPIEN 165  
DB 128 VALSCLKPLPPFHYMTASFGSEV-----PCAS---YRVFGSDALAYEVV-- 170  
QY 166 TPBEKDKERMAHMEYDPCAVLRHGGVYVWGETWEKAKTWCE 211  
DB 171 -----RANGHHR---ACIMASHGMVWVWRDLAHLALAE 201

RESULT 11  
A:Title: ribulose-phosphate 4-epimerase homolog (EC 5.1.3.-) yias - Escherichia coli (strain K-12)  
A:Reference number: A47804; A65158  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-231 <PLU>  
A:Cross-references: EMBL:U00039; NID:g46582; PIDN:AAB18560.1; PID:g466721  
A:Experimental source: strain K-12, substrain MG1655  
A:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-231 <BLAT>

RESULT 13

97977  
probable sugar isomerase (EC 5.1.3.4) [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C/Species: Salmonella enterica subsp. enterica serovar Typhimurium  
A/Note: this species has also been called Salmonella typhimurium  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: A10977  
R/Farkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar AB0502; MUID:21534947; PMID:11677608  
A/Reference number: AB0502; MUID:21534947; PMID:11677608  
A/Accession: A10977  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-231 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD07949.1; PID:gl6504938; GSPDB:GN00176  
C/Genetics:  
C/Superfamily: L-ribulose-phosphate 4-epimerase  
C/Keywords: isomerase

Query Match 10.0%; Score 132; DB 2; Length 231;  
Best Local Similarity 28.5%; Pred. No. 0.00023;  
Matches 61; Conservative 23; Mismatches 94; Indels 36; Gaps 11;

QY 38 GWTGTGGTSLKHEG--IYIAPSGVOKERIQPEDMFVCDINEQDISGPPASKKLKKSQ 95  
DB 20 GLVFTWGNVSVDETKLVKIKSGVEYEWTDWVVEI---ASGKVVGNKKPSS 75  
QY 96 CTPLFMNAY-TMRGAGAVITHSKAAVM---ATLFPQGEFKITHQEMTKGIRKCTS--- 148  
DB 76 DTATHALYRRYPQIGGIVTHSRHATWSQAGLDPA-WGTHADYFYGAIPCTRLMT 133  
QY 149 ----GGYRVDMLVPIENTPEKDKERMAHAMNEYDPSCAVLRHGVVVG--- 200  
DB 134 VEEINGEYEQ---TGEVITKTFEERGLDPA-----QIP---AVLVSHSGPFAWGNAA 181  
QY 201 ETWEKAKTMCEC-YDYLFDAVSMKKMGDLDTQL 233  
DB 182 DAVHNVALEECAYWGLFSRQLAPQDPMPQLL 215

RESULT 14

S73671  
L-ribulose-phosphate 4-epimerase (EC 5.1.3.4) arad - Mycoplasma pneumoniae (strain ATCC 16266)  
N/Alternate names: hypothetical protein P02\_orf242  
C/Species: Mycoplasma pneumoniae  
A/Variety: ATCC 29342  
C/Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
C/Accession: S73671  
R/Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.  
Nucleic Acids Res. 24, 4420-4449, 1996  
A/Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A/Reference number: S73327; MUID:97105885; PMID:8948633  
A/Accession: S73671  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-242 <HM>  
A/Cross-references: EMBL:AE000032; GB:U00089; NID:gl674011; PIDN:AA95992.1; PID:gl67402  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C/Genetics:  
A/Genetic code: SGC3  
C/Superfamily: L-ribulose-phosphate 4-epimerase  
C/Keywords: isomerase; zinc  
F:80,99,101,175/Binding site: zinc (Asp, His, His, His) #status predicted

Query Match 10.0%; Score 132; DB 2; Length 242;  
Best Local Similarity 27.9%; Pred. No. 0.00025;  
Matches 55; Conservative 25; Mismatches 73; Indels 44; Gaps 10;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:29:16 ; Search time 46 Seconds  
(without alignments)  
1653.041 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQDCCSRFCGAQDK.....SMKXGILDPTQLPVGENGIV 241

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1317	100.0	241	11 Q9WVQ5	Q9WVQ5 mus musculus
2	1307	99.2	241	11 Q8BP46	Q8BP46 mus musculus
3	1246.5	94.6	242	4 Q96GX9	Q96GX9 homo sapien
4	1239.5	94.1	242	4 Q9V318	Q9V318 homo sapien
5	1235.5	93.8	242	4 Q8WVU2	Q8WVU2 homo sapien
6	1234.5	93.7	242	4 Q96HK2	Q96HK2 homo sapien
7	786	59.7	227	5 Q9VTF93	Q9VTF93 drosophila
8	772.5	58.7	153	4 Q9H528	Q9H528 homo sapien
9	540	41.0	507	10 Q9FN41	Q9FN41 arabidopsis
10	433.5	32.9	284	5 Q23261	Q23261 caenorhabdi
11	417	31.7	221	3 Q9HE08	Q9HE08 schizosacch
12	289	21.9	240	5 Q8TR31	Q8TR31 heterodera
13	274.5	20.8	192	3 Q9UT22	Q9UT22 schizosacch
14	219.5	16.7	209	16 Q31668	Q31668 bacillus su
15	211	16.0	207	16 Q7W8Y6	Q7W8Y6 prochloroco
16	203	15.4	202	16 Q828L0	Q828L0 streptomyce

17	202	15.3	212	16	Q81956	Q81956 bacillus ce
18	201.5	15.3	211	16	Q7U4V0	Q7U4V0 synechococc
19	201	15.3	212	16	Q81M70	Q81M70 bacillus an
20	195	14.8	204	16	Q884P3	Q884P3 pseudomonas
21	179.5	13.6	249	16	Q8EXC1	Q8EXC1 leptospira
22	166	12.6	217	16	Q8P9N3	Q8P9N3 xanthomonas
23	165	12.5	217	16	Q8PLG0	Q8PLG0 xanthomonas
24	164	12.5	205	16	Q91342	Q91342 pseudomonas
25	161.5	12.3	241	16	Q829J0	Q829J0 streptomyce
26	155	11.8	218	16	Q87C38	Q87C38 xylella fas
27	153	11.6	218	16	Q9PBD5	Q9PBD5 xylella fas
28	148.5	11.3	254	16	Q9WYB9	Q9WYB9 thermotoga
29	146	11.1	231	16	Q8FC7	Q8FC7 escherichia
30	145	11.0	430	16	Q7UYE2	Q7UYE2 rhodospirill
31	144	10.9	190	17	Q8TV16	Q8TV16 methanopyru
32	140	10.6	219	16	Q83HE2	Q83HE2 tropheryma
33	140	10.6	244	16	Q8YDI7	Q8YDI7 brucella me
34	139.5	10.6	240	16	Q9RJ24	Q9RJ24 streptomyce
35	138.5	10.5	191	17	Q27457	Q27457 methanobact
36	136	10.3	226	16	Q8PEU7	Q8PEU7 xanthomonas
37	134.5	10.2	219	2	Q8VQM1	Q8VQM1 brucella ab
38	132	10.0	231	16	Q8Z2D2	Q8Z2D2 salmonella
39	132	10.0	243	16	Q8EPX0	Q8EPX0 mycoplasma
40	131.5	10.0	219	16	Q8FUV6	Q8FUV6 brucella su
41	131	9.9	231	16	Q8ZL75	Q8ZL75 salmonella
42	131	9.9	238	2	Q93Q57	Q93Q57 klebsiella
43	130.5	9.9	236	16	Q93DA6	Q93DA6 streptococc
44	130	9.9	227	16	Q97NJ4	Q97NJ4 streptococc
45	128	9.7	223	16	Q83FR4	Q83FR4 tropheryma

#### ALIGNMENTS

#### RESULT 1

Q9WVQ5 PRELIMINARY; PRT; 241 AA.  
AC Q9WVQ5; DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE MMRP19 (Monocyte macrophage 19).  
GN MMRP19 OR MMR19.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sha S., Aoki Y., Nishi Y.;  
RT "A cDNA sequence from murine monocyte-macrophage."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary Gland;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB028863; BAA78906.1; -  
DR EMBL; BC028434; AAH28434.1; -  
DR MGD; MGI:1926788; Mmrpl9.  
DR InterPro; IPR001303; Aldolase\_II\_N.  
DR Pfam; PF00596; Aldolase\_II; 1.  
SQ SEQUENCE 241 AA; 4366CF4AD2239DB8 CRC64;

Query Match 100.0%; Score 1317; DB 11; Length 241;

Best Local Similarity 100.0%; Pred. No. 2.3e-118;

Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGCQAQDCCSRFCGAQDKHEPRFLIPELCKQFVHLGWVTCTGGGSIKHGNEIYIAPS 60

Db 1 MSGCQAQDCCSRFCGAQDKHEPRFLIPELCKQFVHLGWVTCTGGGSIKHGNEIYIAPS 60

QY 61 GVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTWRGAGAVIHTHSKAA 120

```

Db 61 GVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTSKAA 120
QY 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
Db 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
QY 181 NEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNGI 240
Db 181 NEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNGI 240
QY 241 V 241
Db 241 V 241

```

## RESULT 2

```

Q8BP46
ID Q8BP46 PRELIMINARY; PRT; 241 AA.
AC Q8BP46;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Monocyte macrophage 19.
GN MMRP19.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK077705; BAC36968.1; -.
DR MGD; MGI:1526788; Mmrp19.
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase_II; 1.
SQ SEQUENCE 241 AA; 26933 MW; 5EC51C07E6F0D656 CRC64;

```

```

Query Match 99.2%; Score 1307; DB 11; Length 241;
Best Local Similarity 99.6%; Pred. No. 2.1e-117;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAPS 60
Db 1 MSGCQAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAPS 60
QY 61 GVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTSKAA 120
Db 61 GVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTSKAA 120
QY 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
Db 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
QY 181 NEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNGI 240
Db 181 NEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNGI 240
QY 241 V 241
Db 241 V 241

```

## RESULT 3

```

Q96GX9
ID Q96GX9 PRELIMINARY; PRT; 242 AA.
AC Q96GX9;

```

```

DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE CGI-29 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009077; AAH09077.1; -.
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase_II; 1.
SQ SEQUENCE 242 AA; 27125 MW; 9B8D5D1435D6775A CRC64;

```

```

Query Match 94.6%; Score 1246.5; DB 4; Length 242;
Best Local Similarity 94.2%; Pred. No. 1.4e-111;
Matches 228; Conservative 9; Mismatches 4; Indels 1; Gaps 1;

QY 1 MSGCQA-QGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAP 59
Db 1 MSGCDAREGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAP 60
QY 60 SGVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTSKA 119
Db 61 SGVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTSKA 120
QY 120 AVMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
Db 121 AVMATLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180
QY 180 MNEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNG 239
Db 181 MNEYPDSCAVLVRHGVVWGTEWKAATMCCEYDYLFDIAVSMKKWGLDPTQLPVGNG 240
QY 240 IV 241
Db 241 IV 242

```

## RESULT 4

```

Q9Y318
ID Q9Y318 PRELIMINARY; PRT; 242 AA.
AC Q9Y318;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE CGI-29 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
RT Caenorhabditis elegans by Comparative Proteomics."
RL Genome Res. 10:703-713 (2000).
DR EMBL; AF132963; AAD27738.1; -.
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase_II; 1.
SQ SEQUENCE 242 AA; 27012 MW; 7B99194024C77D5B CRC64;

```

```

Query Match 94.1%; Score 1239.5; DB 4; Length 242;
Best Local Similarity 93.8%; Pred. No. 6.6e-111;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

```

```

QY 1 MSGCQA-QGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAP 59

```

Db 1 MSGCDAGEDCCRRGCAQDKHPRYLIPELCKQFYHLGWVTGTGGISLKHGDEIYIAP 60  
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 119  
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 120  
 QY 120 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
 Db 121 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
 QY 180 MNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 239  
 Db 181 MNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 240  
 QY 240 IV 241  
 Db 241 IV 242

## RESULT 5

Q8WVU2 PRELIMINARY; PRT; 242 AA.  
 AC Q8WVU2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC017594; AAH17594.1; -.  
 DR InterPro; IPR001303; Aldolase\_II\_N.  
 DR Pfam; PF00596; Aldolase\_II; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 242 AA; 27215 MW; F8C4A371E3C63FC2 CRC64;

Query Match 93.8%; Score 1235.5; DB 4; Length 242;  
 Best Local Similarity 93.8%; Pred. No. 1.6e-110;  
 Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCQA-QGDCSRPCGAQDKHPRYLIPELCKQFYHLGWVTGTGGISLKHGDEIYIAP 59  
 Db 1 MSGCDAGEDCCRRGCAQDKHPRYLIPELCKQFYHLGWVTGTGGISLKHGDEIYIAP 60  
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 119  
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 120  
 QY 120 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
 Db 121 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
 QY 180 MNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 239  
 Db 181 MNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 240  
 QY 240 IV 241  
 Db 241 IV 242

## RESULT 6

Q96HK2 PRELIMINARY; PRT; 242 AA.  
 AC Q96HK2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Similar to CGI-29 protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Bone marrow;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC008440; AAH08440.1; -.  
 DR InterPro; IPR001303; Aldolase\_II\_N.  
 DR Pfam; PF00596; Aldolase\_II; 1.  
 SQ SEQUENCE 242 AA; 27112 MW; 691C60447B723C1B CRC64;

Query Match 93.7%; Score 1234.5; DB 4; Length 242;  
 Best Local Similarity 93.4%; Pred. No. 2e-110; Mismatches 10; Indels 1; Gaps 1;  
 Matches 226; Conservative 10; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCQA-QGDCSRPCGAQDKHPRYLIPELCKQFYHLGWVTGTGGISLKHGDEIYIAP 59  
 Db 1 MSGCDAGEDCCRRGCAQDKHPRYLIPELCKQFYHLGWVTGTGGISLKHGDEIYIAP 60  
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 119  
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNAYTMRGAGAVIHTSKA 120  
 QY 120 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179  
 Db 121 AVMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180  
 QY 180 MNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 239  
 Db 181 VNEYDPDSCAVLRHGGVYVGTWEKATMCCEYDLYFDIAVSMKKMGDPTQLPVGNG 240  
 QY 240 IV 241  
 Db 241 IV 242

RESULT 7  
 Q9VY93 PRELIMINARY; PRT; 227 AA.  
 ID Q9VY93;  
 AC Q9VY93;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE CG11134 protein (RE61999p).  
 GN CG11134.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Berkely;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,



RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garb N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarini M., Guarnier B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunco J., Paclob J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Mao K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB003493; AAF48310.1; -  
 DR EMBL; AY071553; AAL49175.1; -  
 DR FlyBase; FBgn0030518; CG11134.  
 DR InterPro; IPR001303; Aldolase II\_N.  
 DR Pfam; PF00596; Aldolase II; 1.  
 SQ SEQUENCE 227 AA; 26011 MW; 7F2E505906CE155D CRC64;

Query Match 59.7%; Score 786; DB 5; Length 227;  
 Best Local Similarity 70.5%; Pred. No. 2.4e-67;  
 Matches 148; Conservative 20; Mismatches 42; Indels 0; Gaps 0;

QY 21 EHPRLPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDINEQ 80  
 DB 12 EHPRLPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDITGK 71

QY 81 DISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSAKAAVMATLLPFGQBFKITHQEMI 140  
 DB 72 DLQLPPEIKGLKKSQCTPLFMAYQHRQAGAVIHTHSAKAAVMATLLPFGKTPRCTHLEMI 131

QY 141 KGIRKCTGGYRYDMLVPIENTPEKDKERMAHMEYPPSCAVLVRHGVYWG 200  
 DB 132 KGVIDEADKRYLRYDEELVPIENTPTPEKDLADSYAAMMEYPCGSAILVRHGVYWG 191

QY 201 ETWEKAKTMCYDYLFDAVSMKMGGLDP 230  
 DB 192 QNWEKAKTMCYDYLFDAVSMKMGGLDP 221

RESULT 8  
 Q9H528 PRELIMINARY; PRT; 153 AA.

AC Q9H528  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE DJ179L10.2 (Similar to CGI-29 protein) (Fragment).  
 GN DJ179L10.2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moore M.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL138810; CAC12642.1; - II\_N.  
 DR InterPro; IPR001303; Aldolase II\_N.  
 DR Pfam; PF00596; Aldolase II; 1.  
 FT NON TER 153  
 SQ SEQUENCE 153 AA; 16918 MW; D88A0ADE331E12EA CRC64;

Query Match 58.7%; Score 772.5; DB 4; Length 153;  
 Best Local Similarity 92.8%; Pred. No. 2.9e-66;  
 Matches 142; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 MSGCOA-QGCCSRPCGAQDKEHPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYAP 59  
 DB 1 MSGCDAREGDCSRRCAQDKEHPRYLIPELCKQFVHLGWVTGTGGISLKHGDEIYAP 60

QY 60 SGVQKRIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSA 119  
 DB 61 SGVQKRIQPEDMFVCDINEKDISGPPSKLKKSQCTPLFMNAYTMRGAGAVIHTHSA 120

QY 120 AVMATLLPFGQBFKITHQEMIKGIRKCTSG3YY 152  
 DB 121 AVMATLLPFGREFKITHQEMIKGIRKCTSG3YY 153

RESULT 9  
 Q9FN41 PRELIMINARY; PRT; 507 AA.

AC Q9FN41  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Similarity to enolase-phosphatase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=98162728; PubMed=9501997;  
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.  
 RT Sequence features of the regions of 1,191,918 bp covered by seventeen  
 RT physically assigned P1 clones.";  
 RL DNA Res. 4:401-414(1997).  
 DR EMBL; AB007644; BAB10715.1; -  
 DR GO; GO:0016787; F-hydrolase activity; IEA.  
 DR GO; GO:0008152; P-metabolism; IEA.  
 DR InterPro; IPR001303; Aldolase II\_N.  
 DR InterPro; IPR005834; Hydrolase.  
 DR Pfam; PF00596; Aldolase II; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 SQ SEQUENCE 507 AA; 56520 MW; 2F99D9F52FA1D5DC CRC64;

Query Match 41.0%; Score 540; DB 10; Length 507;  
 Best Local Similarity 49.6%; Pred. No. 2.8e-43;  
 Matches 112; Conservative 32; Mismatches 60; Indels 22; Gaps 7;

QY 26 LIPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDI 77  
 DB 27 LVTELCHRFYTGQWVSGTGGISITMKVHDASIPKPEQLIVMSFGVQKERMQEDMYILSA 86

QY 78 NEQDISGPP-----ASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSAKAAVMATLLP-GEFF 132  
 DB 87 NGSIISTSPKPYNPKPKCTDCAPLFMKAYEMRNAGAVIHTSHGMSCLVTMLNPQAEF 146

```
QY 133 KITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 192
DB 147 RITHMEMIKGIQ---GHGY---DELAVPIIENAYENELTSLTKALEAYPKATAVLVR 200
QY 193 RHGVVWGTETWEKATKMECEVDYLDIAVSKMKGLD---PTQLPV 235
DB 201 NHGVYWGDSWTHAKTQAEVHYLFDAAIKHLQGLDAATPDHGPI 246

RESULT 10
Q23261 PRELIMINARY; PRT; 284 AA.
AC Q23261;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE ZC373.5 protein.
GN ZC373.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]_
RN [2]_
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [1]_
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z49131; CA88977.1; -.
DR FIR; T27523; T27523.
DR WormPep; ZC373.5; CE02378.
DR InterPro; IPR001303; Aldolase II_N.
DR Pfam; PF00596; Aldolase II; 1_
SQ SEQUENCE 284 AA; 32367 MW; 4DC5B94439F2E529 CRC64;

Query Match 32.9%; Score 433.5; DB 5; Length 284;
Best Local Similarity 39.9%; Pred. No. 2.4e-33;
Matches 87; Conservative 35; Mismatches 75; Indels 21; Gaps 1;

QY 29 ELCKQFYHLGWGTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEODISGPPAS 88
DB 35 ELMIQFYKLGWRGSGGANGCTISGSELMISPSALQKRIREQDVFYNNKDKTEVQRPN 94
QY 89 KKLKKSQCTPLFMNAYTMRGAGAVITHSKAAVMATLLPFGQEFKITHQEMIKGIRKCTS 148
DB 95 KRITVSSCSVLFLIMKETGSECVITHSKCANLITOLIKSNVFEISHQEIYKGIYDPFS 154
QY 149 GGYRYDDMLVPIIENTPPEEKDL-----KERMAHAMNEYDPDSC 187
DB 155 GKALKYSDTLTPIIDNMPSESQLLVCALNRSPMQPDCTINIALFOEPIRGVLENTPOAI 214
QY 188 AVLVRHGVVWGTETWEKATKMECEVDYLDIAVSKMK 225
DB 215 AVLVRHGVWGTWESTKINTECIDYLLLELSIEMLK 252

RESULT 11
Q9HE08 PRELIMINARY; PRT; 221 AA.
AC Q9HE08;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative phosphate epimerase.
GN SPAC20H4.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
```

```
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Seger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL512487; CAC19735.1; -.
DR GeneDB SPombe; SPAC20H4.05C; -.
DR InterPro; IPR001303; Aldolase II_N.
DR Pfam; PF00596; Aldolase II; 1_
SQ SEQUENCE 221 AA; 25195 MW; B959011B581EAB74 CRC64;

Query Match 31.7%; Score 417; DB 3; Length 221;
Best Local Similarity 39.6%; Pred. No. 6.8e-32;
Matches 90; Conservative 40; Mismatches 75; Indels 22; Gaps 7;

QY 3 GCOAGDCCSRPCGAQDKHEPRFLIPELCKQFYHLGWGTGGISLKHGNEIYIAPSGV 62
DB 10 GCLRSGLD--KXCGE-----LICEICRDLYTSGWVTG-----GDAIVIAPSGV 51
QY 63 QKRIQPEDMFVCDINEODISGPPASKKLKKSQCTPLFMNAYTMRGAGAVITHSKAAV 121
DB 52 QKERMELHFLFVMSLITREYMPMPA-LRLKPSQCTPLFLAVVTSLRDAVACIHTSQEAI 110
QY 122 MATLLPFGQE-FKITHQEMIKGIRK-CTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHA 179
DB 111 LLSTLTFADSDHFSATGFEVLSYIPKSGKNNGPHKPTDKIKIPFINNTAHESDLHDSLOEA 170
QY 183 MNEYDPDSCAVLVRHGVWGTETWEKATKMECEVDYLDIAVSKMK 226
DB 171 INLYPDTCAVIRDHGIYCWGDTQDWTQNTAEVSLFQAYLRRRL 217

RESULT 12
Q8TA31 PRELIMINARY; PRT; 240 AA.
AC Q8TA31;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Heterodera glycines (Soybean cyst nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
OC NCBI_TaxID=51029;
RN [1]_
RP SEQUENCE FROM N.A.
RA Gao B., Allen R., Baum T.J., Davis E.L., Hussey R.S.;
RT "Cloning putative parasitism genes from Heterodera glycines.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF345795; AAL78223.1; -.
DR GO; GO:0005874; C:microtubule; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007018; P:microtubule-based movement; IEA.
DR InterPro; IPR001303; Aldolase II_N.
DR InterPro; IPR002453; Beta tubulin.
DR Pfam; PF00596; Aldolase II; 1.
DR PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 240
SQ SEQUENCE 240 AA; 26062 MW; 77862AEE08142DBA CRC64;

Query Match 21.9%; Score 289; DB 5; Length 240;
Best Local Similarity 35.1%; Pred. No. 1.5e-19;
Matches 67; Conservative 37; Mismatches 77; Indels 10; Gaps 4;

QY 27 IPCLCKQFYHLGWGTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEODISGPP 86
DB 17 LAELIRHFALGWRDNGGMAVLNGAVFGSPTSQKEKVPENDLFDVDTATGTVLKRP 76
QY 87 ASKKLKKSQCTPLFMNAYTMRGAGAVITHSKAAVMATLLPFGQEFKITHQEMIKGIRK 146
```

Db 77 QNAAVSPTSATGGLMNT-GLNCVITHSKYANLVSQLVGTGNEFAIQNQMIGQYENR 132  
Qy 147 TSGGYRYD-DMLAVPIENTPEKDKERMAHANEYPDSCAVLVRHGVYVWGGETWE 204  
Db 133 SSG--LRDNDVRUVFVDSSELNEQMLSPVLLRTDKYTEASALLVRGHGFFVFGSSLA 190  
Qy 205 KA--KTMCECY 213  
Db 191 GSLPKWMLF 201  
RESULT 13  
Q9UT22 PRELIMINARY; PRT; 192 AA.  
AC Q9UT22;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Conserved hypothetical protein.  
GN SPAC9.06C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h;  
RA Wedler H., Dusterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL121764; CAB57424.1; -;  
DR PIR; T39191; T39191.  
DR GeneDB SPombe; SPAC9.06C; -;  
DR InterPro; IPR001303; Aldolase\_II\_N.  
DR Pfam; PF00596; Aldolase\_II; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 192 AA; 21797 MW; F988CD3983007191 CRC64;  
Query Match 20.8%; Score 274.5; DB 3; Length 192;  
Best Local Similarity 34.6%; Pred. No. 2.8e-18;  
Matches 71; Conservative 36; Mismatches 73; Indels 25; Gaps 7;  
Qy 26 LIPELCKQFYHLGWVT-GTGGGSLKHGNEIYAPSGVQKERIQPEDMFVCDINEQDISG 84  
Db 9 LIELIPIHYSLGWKKFGSGYALCVK-----DRVQRDFITENDIVTFNLSQSVT- 58  
Qy 85 PPASKKLKKSQCTPLFMNAYTMRGAGAV--IHTHKAAMVATLLFPQGEFKITH--QEM 142  
Db 59 -----KDLVNWAYIFSWLSNMDAVACIYSTSVAAVGASMY--NEKFTTQSKEMING 108  
Qy 143 IRKCT-SGGYRYDDMLVPIENTPEEKDKERMAHANEYPDSCAVLVRHGVYVWGE 201  
Db 109 IPKGNPSAGYLCFFDTLEVPITHN-GDSKTIIDELKKVIELPQICAVLIRGHGVWGGA 167  
Qy 202 TWEKAKTMCECYDYLFDIAVSMKKM 226  
Db 168 TWEKAKTMCECYEYLFELDYKLKTL 192  
RESULT 14  
O31668 PRELIMINARY; PRT; 209 AA.  
AC O31668;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE YKRY protein.  
GN YKRY.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerston I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoeft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Rapoport G., Rey M., Reynolds S.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rose M., Sadale Y.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z99111; CAB13234.1; -;  
DR PIR; A69864; A69864.  
DR InterPro; IPR001303; Aldolase\_II\_N.  
DR Pfam; PF00596; Aldolase\_II; 1.  
KW Complete proteome.  
SQ SEQUENCE 209 AA; 23489 MW; D4E8BE3544026FAB CRC64;  
Query Match 16.7%; Score 219.5; DB 16; Length 209;  
Best Local Similarity 30.9%; Pred. No. 6.1e-13;  
Matches 64; Conservative 34; Mismatches 88; Indels 21; Gaps 7;  
Qy 27 IPELCKQFYHLGWVTGTGGGSLKHGNE--IYIAPSGVQKERIQPEDMFVCDINEQDIS 83  
Db 11 LAEVRKELAEKRDWFPATSGNLSIKVTDEPLFLVTSAGDKRKETVEDFLVDQN----- 65  
Qy 84 GPPASK--KLKKSQCTPLFMNAYTMRGAGAVIHTHKAAMVATLLFPQGEFKITH--QEM 139  
Db 66 GEPAESGHSKPSAETLLHLYNKTNAGCCLHVTNNVISELY-GDQKITPKGQEI 124  
Qy 140 IKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHANEYPDSCAVLVRHGVYVW 199  
Db 125 IKAL-----GLWEENAETVPIIENPAHIPTLAALFAEIESE--DSGAVLIRNHGITAW 176  
Qy 200 GETWEKAKTMCECYDYLFDIAVSMKKM 226  
Db 177 GKTAFAEKRVLEAVEYFLFSYHLKLT 203  
RESULT 15  
Q7V8Y6 PRELIMINARY; PRT; 207 AA.  
ID Q7V8Y6  
AC Q7V8Y6;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

